

China's misconception of eugenics

China's plans for eugenics must be judged by the degree to which they interfere with people's wishes; they may not differ much from programmes followed elsewhere but compulsion will make them unacceptable.

THERE are several reasons why the concept of eugenics deserves a bad press. For one thing, the concept has discreditable associations — Adolf Hitler, the slaughter of the congenitally disadvantaged (some with genetically heritable diseases) and, more wrong-headedly still, the expressed belief that the pursuit of genetic homogeneity in a supposed Aryan race of people would be at once feasible, desirable and advantageous. More generally, the concept has often been a cloak for the illiberal yearnings of supposedly liberal peoples (such as those of Europe earlier this century) for a better, and usually a long-vanished, way of conducting social affairs.

Of course, most modern governments hope that their people will be energetic, ingenious and enterprising. (The British prime minister, Mr John Major, was speaking on that theme only the other day.) Usually, they seek to bring about that state of affairs by social engineering of one kind or another, and by their provision of education in particular. Now the government of China, which has never been conspicuous among the ranks of the politically correct, has gone a big step further with its draft law on eugenics (see page 3).

First reactions to these proposals are likely to be hostile, but that is probably too hasty a judgement. It is not yet clear how the Chinese government intends that its new law should be enforced. If the intention was merely that people in China should be warned of the risks of inheriting heritable genetic diseases, and helped to avoid them among their children, the new law would not be very different in its effect from the services provided elsewhere, where public (and otherwise private) health services offer genetic counselling, amniocentesis and even, on occasion, abortion if there is proof that the outcome of a pregnancy will be a seriously handicapped child. The chief beneficiaries of those services, which are of necessity voluntary, are parents and their children. To the extent that seriously handicapped children may be (and in the last resort, should be) a charge on the public finances, there may also be some benefit to national exchequers, but that can be only marginal.

The Chinese proposal that those with a medical history of undesirable traits be forbidden to marry suggests that the compulsory element may be larger than desirable. But the draft law, as disclosed, at least does not follow the Third Reich in offering the goal of some genetic ideal. Rather, it is concerned with negative eugenics, or the avoidance of avoidable genetic handicap among future generations. That

is as it should be. But the proposals are also shot through with inconsistency and ambiguity suggesting that, if they were rigorously enforced, they might be more damaging than beneficial. One curious feature of the new law, for example, is that it lumps together genetic diseases with "mental illness" and hepatitis as counter-indications of pregnancy or even marriage.

Nobody denies that hepatitis-B infection is a serious public health problem in China, especially in the coastal regions, but what purpose can be served by avoiding the birth of children to parents who are, or who have been, infected? If infection is taken as a sign that the progeny will also be infected, a public health programme (long overdue, in any case) would be the more seemly remedy. If hepatitis-B infection is being taken as a proxy for an undesirable lifestyle, that would be illiberal.

The "mental illness" disqualification is a still more serious minefield. The experience of the former Soviet Union is that the practice of psychiatry can easily become an extension of the penal system, a means of expressing social (or at least official) disapproval. It will be remarkable if China remains perpetually free from the temptation to follow suit (even if the chances of finding a gene for political dissent are slim). But too little is known of the genetics of psychiatric illness (and even of more tangible neurological disorders) for people to be given accurate genetic counselling. The mind boggles at the prospect of counselling people with a family history of Huntington's disease, where the chance that a child will develop the disease appears to depend on the chance that a triplet of nucleotides in the gene will have expanded beyond some critical point. And what if, say, the gene 'for' manic-depressive illness (if there is one) is recessive and if the heterozygous advantage it confers is just the entrepreneurship on which present-day China seems to have pinned its faith in the future?

The truth, sadly for China's eugenic ambitions, is that even the present (and much enlarged) understanding of genetics does not permit the eradication of genes supposed to be harmful except by searching in the genome of every putative fetus for the DNA characteristic of such conditions as are known. Worse, there is every likelihood that gene products are generally multivalent, performing essential tasks in some organs, or at some stages in the life cycle, while presenting risks to (or at) others. Otherwise, it is reasonable to ask, why has not natural selection got rid of them?



In other words, China's eugenic ambitions are no more likely to be attainable now than they would have been at the beginning of this century, when so much less was known. In the end, like other countries, China will have to learn to live with the social and economic burden of disability apparently unavoidable in a diverse community. In the meantime, it is earnestly to be hoped that misplaced ambition does not tempt the Chinese government to impose ineffectual constraints on its vast population, still reeling from the draconian aspects of the "One couple, one child" regime. □

Too old at 59?

The birth of artificially conceived twins to a woman of 59 has provoked a characteristically British fuss.

FOR some years there has been a sprinkling of births to postmenopausal women, all born of the recognition that the chief characteristic of the human menopause is the disappearance of the hormones that make conception and pregnancy possible. The first cases were publicly reported from California three years ago, but last week the British newspapers were telling, with customary sanctimonious humbug, how a British woman had given birth to twins at the ripe old age of 59.

Proper anxieties about the continuing war in Bosnia, the political succession in Russia and the continuing high level of unemployment in Europe notwithstanding, several politicians have discerned further evidence of social disintegration in what would traditionally have been regarded as a happy event. The general opinion is that "something will have to be done". Mrs Virginia Bottomley, Secretary of State for Health, on this occasion more circumspect than usual, followed the remark that "women do not have a right to have children" with a promise that she would bring up the ethical issues raised by the artificial conception with her counterparts in the other 11 governments of the European Union (the prenatal procedures had been carried out in Rome).

The difficulty is that no new ethical issues are raised by the application of *in vitro* fertilization (IVF) to older women. As now practised, IVF is controlled by physicians, whose first responsibility must be to their patients and includes the duty to counsel them about the risks involved in giving birth at an unusually great age. Plainly there are several: not merely the increased risks of childbirth itself, but the risk that it will not be possible to provide a child with the personal care it needs during its formative years. The same questions arise with women seeking more common IVF treatment, for they are at the older end of the usual spectrum of child-bearing age. At the very least they have usually been seeking pregnancy by natural means for some years before going to an infertility clinic.

So there is no sharp division between the ages of the women seeking help nor is the treatment offered sharply different. Bottomley is thus likely to be treated to indulgent smiles when she brings up the matter with her less sheltered European colleagues, which is just as well. □

What priority for TB?

Western governments must recognize the need to combat TB in developing nations — if only in self-defence.

THE World Health Organization (WHO) recently published a report on tuberculosis with a dramatic title that neatly sums up the reaction of governments everywhere to what is now the most prevalent fatal infection: *TB: A Global Emergency — Low Priority*. And that is exactly the spirit in which the US Agency for International Development (AID) recently turned down a request from WHO for \$3 million to fight TB in developing countries. (As part of a United Nations campaign, WHO is trying to raise a total of \$9 million.) Yet tuberculosis is likely to kill 30 million people before the turn of the century. The number could be higher if drug-resistant strains of the TB mycobacterium continue to proliferate.

But US AID, a \$6.5 billion agency, could not come up with the \$3 million: Ann Van Dusen, an AID officer, said: "In the budget environment that we're looking at, it is a lot of money". That is either just plain nonsense (\$3 million is not a lot of money for the US government) or else represents priorities distorted beyond belief. In a news report, Van Dusen defended the agency's decision, saying that AID is already spending a lot of money on AIDS, for instance.

And an AID officer correctly pointed out that the agency has money for TB in many of its local health programmes. But no total dollar sum is available, which in itself is a sign of TB's low visibility within AID.

Does AID not realize the lethal connection between tuberculosis and AIDS? Not only in developing countries but also in Western nations, TB is flourishing because AIDS patients whose immune systems are severely compromised are ideal hosts. It is regrettable but true that AIDS patients provide the bacillus with a ideal reservoir in which to mutate into drug-resistant forms. The best of all strategies, therefore, would be to tie TB and AIDS fundings together whenever possible. After all, to prevent the spread of drug-resistant TB in developing nations is to protect the populations of more developed nations as well.

Unlike AIDS, tuberculosis is not difficult to catch. A readily transmissible disease that has even been passed on to passengers on airplanes, it is urgent to take its resurgence seriously. According to WHO data, US AID is not the only national agency to resist the call for funding in the fight against TB. European nations too have been reluctant to contribute.

This sets the stage for a global tragedy. Unlike AIDS, cancer, neurodegenerative diseases and other maladies that await new scientific understanding before they can be cured, the treatment of TB (particularly non-resistant strains) is well-understood and in developing nations is relatively inexpensive—as little as \$30 a patient.

If ever there were a challenge worth taking on, the fight against the worldwide spread of tuberculosis should top the list. World governments should do all they can to negate WHO's statement that TB is being rated "low priority". □

Concern grows over China's plans to reduce number of 'inferior births'

London. Plans announced by the Chinese government last month to introduce a new law designed to eliminate "inferior births", for example by forbidding those with a hereditary disease from marrying, have provoked widespread concern in the international scientific community.

There is general sympathy among scientists with the problems faced by China due to the continuing growth of its population. There is also some acknowledgement that certain measures generally disapproved of by the West — such as limiting families to a single child — may be more acceptable in an environment where social and economic problems are acute.

China's own scientific community has in general backed the government's efforts to impose a strict population control policy. But some Western scientists and genetics experts complain that the new measures designed to eliminate "inferior births" could lead to the spectre of a Nazi-style eugenics programme.

The new policy is contained in a draft law on "eugenics and health protection" presented on 20 December to the Eighth National People's Congress (NPC) Standing Committee in Beijing by the Minister of Public Health, Chen Minzhang. Under the proposed law, those with ailments such as hepatitis, venereal disease or mental illness "which can be passed on through birth" will be banned from marrying.

The cost of looking after those with hereditary handicaps was enormous, said Chen, imposing a heavy burden both on the state and on millions of families. There was therefore wide popular support, he added, for the rapid enactment of a eugenics law containing effective measures to reduce inferior-quality births.

Western observers say that the principles behind China's new proposals — for example, suggesting to pregnant women diagnosed as carrying an abnormal fetus that she might consider an abortion — are little different from those practised by genetic counsellors in the West.

But the fear that, by introducing a compulsory element, China may be embarking on a programme of involuntary "genetic cleansing" has generated considerable heat. Sir Francis Graham-Smith, for example, of the Royal Society, one of the organizers of a meeting on population held by the world's scientific academies in New Delhi in November (see *Nature*, 366, 3; 1993), says he is "horrified" at various aspects of China's plans.

Some critics have focused on the criteria

the Chinese government may use to identify those selected as the targets of recommended abortions or sterilization, or enforced marriage bans. Chen, for example, has said that births of "inferior quality" are serious among "the old revolutionary base, ethnic minorities, the frontier and economically poor areas".

"This policy is being guided by pseudo-science," says Xin-Yuan Fu, a biochemist at Mount Sinai School of Medicine in New York and a member of the human rights committee of the New York Academy of Sciences.

Others add that suspicions of a political motivation are reinforced by the fact that attempts to "breed out" unwanted genetic characteristics are likely to be ineffective, partly because of the difficulty in identifying and controlling carriers in the population (short of mass screening and steril-

ization campaigns).

Lurking behind these concerns is fear of what could happen if China's new policy becomes linked to advanced genetic screening techniques, and that this could stimulate a backlash against these techniques in the West. China is already developing skills in both genome sequencing (see *Nature*, 365, 22; 1993) and in gene therapy, claiming for example to have successfully treated patients with haemophilia in this way.

Western scientists have welcomed the potential contribution of these techniques to China's medical problems. But they are worried about possible consequences if such techniques are applied without sufficient caution, claiming that China has no legally enforceable ethical guidelines in the field of genetic screening and gene therapy.

David Dickson

Astronomer to head new research council

London. The British government has moved to reassure physicists about its concern for the future of basic science by appointing a widely-respected astronomer to run the new Particle Physics and Astronomy Research Council.

Kenneth Pounds, at present professor of space physics at the University of Leicester, will take over the new post of chief executive at the beginning of April. This is when the existing Science and Engineering Research Council is to be split up, with the bulk of its activities being divided between PPARC and a second new body, the Engineering and Physical Sciences Research Council (EPSRC).

Pounds is well known as a pioneer of X-ray astronomy, and has played a central role in a number of recent space missions involving X-ray satellites. He has also been involved in exploring with industry the broader use of detectors developed for these satellites, for example in the analysis

of dental X-rays; this experience appears to have been one factor which persuaded the government to invite him to apply for his new post.

Pounds says that his first challenge will be to get physicists and astronomers work together. "These two strong-minded communities have not been natural bedfellows in the past, but now we are going to sink or swim together," he says.

The other problem he faces will be funding. "I have not seen the books, but I would be surprised if the financial situation is not very tight," says Pounds, expressing particular concern for the balance between the high level of subscriptions to international facilities, and the relatively low spending on domestic efforts needed to take advantage of them.

At the same time Pounds says he would like to see Britain making any early commitment to the construction of the Large Hadron Collider at CERN. "To defer a decision on the LHC would be a mistake. These programmes have an optimal timescale, and the sooner we get committed to this programme, the earlier we will be able to start talking turkey to the Americans and anyone else. If you want to lead an international programme, to must first commit to it, and then go out and seek partners." □



Kenneth Pounds

ESO countries bury hatchet over new optical telescope

Munich. Sharp differences over the plans and costs of what will eventually be the world's largest optical telescope, the Very Large Telescope (VLT), to be built at Mount Paranal in northern Chile, have been resolved by the member states of the European Southern Observatory.

As a result, construction work on the infrastructure for the VLT will begin as planned next month. Installation of the first of the four main 8.2-metre telescopes making up the core of the VLT is due for completion in late 1996 or early 1997, and of all four by 2000.

But plans have been postponed to link the four telescopes into the VLT Interferometer (VLTi). Also delayed will be installation of three smaller outlying telescopes and their associated instrumentation which, together with the original telescopes, will form VLT Interferometer Sub-Array (VISA).

Under original plans approved in 1987, all seven telescopes were to have been built as a single project, and the whole system would have come into operation in 2001. Since then, however, tension has grown over an anticipated 25 per cent increase in its estimated costs over the initial figure of DM463 million (US\$275 million).

One reason for the increase has been demands from astronomers for increasingly sophisticated equipment, such as the 'adaptive optics', a French innovation allowing a small fast-moving mirror to compensate for distortions caused by air turbulence.

The additional costs had led to growing complaints from several of the eight countries which contribute to ESO. France and Italy were particularly vocal. Indeed France recently argued that, in order to keep within original estimates,

the number of main telescopes should be reduced to three, or perhaps even two.

In the light of these complaints, a new construction schedule was agreed unanimously by the ESO council at the beginning of December. This will allow the VLT itself to begin observation by early 1997; indeed the opportunity to concentrate on a narrower range of targets may allow the date to be pulled forward by a few months.



Artist's model of the Very Large Telescope (VLT)

Work will also begin on foundations for the other equipment. But no industrial contracts will be signed for this equipment for at least another two years, and perhaps longer.

Stretching out the overall construction means that the ESO's budget contributions from its member states next year will fall from \$142 million (US\$84 million) as planned to DM123 million.

Experiments with the complete system will now be delayed until well into the next decade, frustrating astronomers who had hoped it would be in operation earlier. But Riccardo Giacconi, the director general of ESO, says he is not unduly worried. "We were trying to do too much in a situation where money is tight," says Giacconi. "We now have a better chance of success."

Alison Abbott

Pay battle threatens fusion research

Oxford. In a bid to end a long-standing dispute over the treatment of British scientists working for the Joint European Torus (JET) in Oxfordshire, the European Parliament has voted to withhold ECU59 million (US\$67 million) from the European commission's fusion budget for 1994 until the dispute is resolved.

The parliament is backing attempts by British staff at JET to receive the same salaries and future job opportunities as those from other European states. The

money being withheld, which represented about 30 per cent of the total fusion budget, will only be released once the dispute is settled.

But there is little prospect of this happening in the near future. A spokesman for AEA Technology, which employs British staff at JET, said that it cannot increase their pay as demanded, as this would fall foul of the Britain's public sector pay policy under which pay rises are limited to 1.5 per cent

Oliver Tickell

US admits to use of humans in radiation experiments

Washington. Administration officials met at the White House this week to coordinate a mammoth inquiry into radiation experiments that have been conducted in the United States since the Second World War.

The search for details on all such experiments will involve a number of government agencies. It is sure to embarrass parts of the scientific community, although many scientists and ethicists said last week that they welcomed the scrutiny.

The interagency initiative follows recent moves by Hazel O'Leary, the Secretary of Energy, to re-examine controversial radiation experiments conducted on human subjects by her own department and its predecessor, the Atomic Energy Commission.

O'Leary has promised to compensate up to 800 known subjects of the experiments and, sensing a public tide in her favour, has opened up a telephone hotline for people claiming to have information about such tests.

Her actions have opened the floodgates to a host of stories — many previously reported to little public response — of experiments conducted in the period 1945-75 which subjected people to varying degrees of radiation, often without their consent.

O'Leary has appointed Dr Ruth Faden of Johns Hopkins University to head an advisory task force to review material from inside the department and from the hotline.

As stories of sometimes gruesome experiments involving children, servicemen, prisoners and the mentally handicapped sweep across America's television screens — many attributed to researchers at top institutions such as Harvard and the Massachusetts Institute of Technology — scientists are deeply divided over the implications of the affair.

Some senior scientists believe that the ethical transgressions in many of the experiments were minor by the standards of the time, and that the whole field of biomedical research has in any case been transformed for the better since the experiments took place.

But others are less sanguine about what they see as the central role of the scientific establishment in allowing the experiments to proceed in the first place. Dr David Egilman, a physician at Brown University, Rhode Island, backs an investigation by the American Public Health Association: "That's the only organization with clean hands," he says.

Colin Macilwain

1994: Science enters the post-Cold War era

The past year has seen continuing transition on the global stage. The storming of the Russian Parliament in October indicated that, whatever the future of the former communist countries of Eastern Europe, a return to the rigid East–West confrontations of the Cold War is unlikely. The successful termination of negotiations on the General Agreement on Trade and Tariffs has established the codes of conduct under which the global wars of the immediate future — those over access to markets — will be fought.

In the following pages, our correspondents around the world report on the impact these trends are already having on the way research is conducted. The most obvious, namely the reduced need for heavy spending on military research, is perhaps the least significant. There has been little evidence that funds freed in this way are being directly transferred to civilian research; indeed it would be surprising to see the so-called ‘peace dividend’ surface in this simplistic way.

More significant has been the need to fill the role previously played by government support of military research in maintaining the health of the science base, particularly in countries such as the United States and the former Soviet Union. It is this, together with a shift in concern from military to economic security, that has helped to determine the new world scientific order emerging in countries of both East and West, North and South.

What are the main characteristics of this new order? First, that governments both conservative (as in Britain, France and Germany) and liberal (as in the United

States, Canada and Australia) feel a new responsibility to back research policies designed to boost ‘strategic’ civilian technologies given the market’s inability to address issues of national security.

Second, that in a world where the ability to prosper depends on the quality

Third, scientific projects whose motivation lay partly in national prestige — itself a valuable currency during the Cold War — have lost much of their lustre. The Superconducting Super Collider (SSC) in the United States is one such casualty. A similar fate could soon meet NASA’s planned ‘international’ space station, while France is under pressure to reduce funding for large projects in fields such as space and nuclear energy. Science as a political spectacle has lost its bite.

As the reports below indicate, the post-Cold War era has brought new promises for science, including a renewed political commitment (valuable when the welfare state is under attack) and a determination to address shortcomings in science and engineering education. But it has also brought new threats, from the marginalization of research whose potential contribution to wealth or well-being cannot be demonstrated, to the shrivelling of support for research teams (for example in the former communist states or in the newly industrialized nations) ill-equipped to survive the harsh winds of international competition.

Scientists may have less to fear than they did at the beginning of 1993; the promise of renewed economic growth can only be good news for those whose funding prospects depend directly on the health of the economy that supports them. But, with research budgets being frozen virtually everywhere, and pressure increasing to evaluate all research projects in terms of competitive advantage, 1994 does not yet promise much cause for celebration. □

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Point of no return: Muscovites watch the attack on the Russian parliament.

of technical skills as judged by the market-place, the criteria used to evaluate research projects are increasingly those accepted by the international scientific community, rather than those that meet local self-interest. Hence the growing willingness of countries such as Japan, France and Italy, previously suspicious of the judgement of outsiders, to open their research systems to external review.

US looks cautiously to Clinton for leadership into new territory

Washington. The opening year of President Bill Clinton’s administration cut both ways for science. The scientific community was able to applaud the appointment of a string of its respected members to posts in the administration. But many also saw the closure of the Superconducting Super Collider (SSC) as an ominous turning point in Washington’s approach to science funding.

This year will see how far these hopes — and fears — are justified. In particular, it will demonstrate the impact of what was perhaps 1993’s most profound development in science policy, namely the creation of a National Science and Technology Council. This ‘science cabinet’ will be chaired by Clinton himself, and will oversee the federal government’s entire \$76-billion research budget.

The appointments of Harold Varmus, the Nobel prize-winning molecular biologist, as director of the National Institutes of Health (NIH), and of the respected physicist Neal Lane to head the National Science Foundation (NSF), reassured the scientific community that the case for basic research would be strongly argued in top political circles.

Such support is likely to become increasingly necessary in a political climate that requires every dollar spent on research to be fully justified in terms of some social outcome, whether it is greater industrial competitiveness or a lower infant mortality rate.

No such justification was forthcoming for the ill-fated SSC. Policy makers are looking to the new council to provide a clear direction for science, to replace the fog in which the Texan atom smasher was lost.

Varmus, for one, describes it as “very useful and potentially very important”; Bruce Alberts, president of the National Academy of Sciences, says the council will work if Clinton wants it to do so.

The will is certainly there. At a recent White House reception for Nobel laureates, Clinton and his wife Hillary seemed to impress their guests. Varmus says they were “informed and enthusiastic” on science policy matters, and predicts the president will chair the council with gusto.

Guided by Jack Gibbons, the president’s science advisor, the administration appears to have accepted the importance of fundamental science. But Congress is likely to continue to press strongly for research that meets “strategic” goals. And the first to come in for close scrutiny will be the NSF.

In a few weeks, Lane is due to appear before the Senate appropriations subcommittee chaired by Barbara Mikulski (Democrat, Maryland) to explain how he plans to make sure that 60 per cent of NSF’s work ►

is 'strategic'. Lane — and the rest of the science community — will either have to convince Congress that much basic research can comfortably wear that label or accept that research directions will have to be changed.

Supporters of science in Congress are sanguine about the outcome. "My feeling is that there's going to be no lessening of funding for basic research, but probably a little more guidance on which areas are critical for the nation — as we did before with defence-related science," says George Brown, chairman of the House Science, Space and Technology Committee.

But if funding for civil science is to grow — and 1993 has seen healthy budget increases at least at the NSF and the National Institute of Standards and Technology (NIST), the administration's chosen workhorse for technology transfer — defence research must continue to contract: there is nowhere else for the money to come from.

The ending of the Cold War would seem to point in this direction. Indeed, efforts are being made to ensure that concern for economic security delivers the same pay-off to the scientific community as did the previous concern for military security.

But the shift cannot be guaranteed, particularly with the recent nomination as Secretary of Defense of Admiral Bobby Inman, who has already expressed his independence from the administration in remarkably forthright terms, and is seen as the dark horse in the new "science cabinet". Inman has previously shown strong interest in research and high technology. A successful rearguard action by him to shore up the vast defence research budget could wreak havoc with Clinton's plans for expansion elsewhere.



Keen listeners: Hillary and Bill Clinton impressed scientific guests at a recent White House reception.

More immediate concerns are likely to dominate the science policy agenda in the coming months. Proposals for health care reform will become sufficiently concrete to influence the future path of biomedical research. And NASA's proposed "international" space station will fly a dangerous (and possibly fatal) mission into Congress.

But the space station's destiny will be of more concern to the aerospace industry than to most scientists. The latter are likely to be more worried about a slow — and Hollywood-assisted — erosion of their public image. The dedicated, white-coated researcher, labouring to save the planet from poverty and disease, risks being supplanted in the public mind by a money-grabbing, plagiarizing con-artist.

This phenomenon has some way to go. A recent opinion poll showed the National Academy of Sciences to be the most credible of all American institutions, well ahead of the military in second place — and with six times the level of public confidence enjoyed by either Congress or the media.

But top scientists have been alerting their peers to the risk of losing this reputation. Alberts and Lane have been among those urging scientists to get out to schools, offices and factories and win the public's appreciation. Outside Washington, that is the priority for 1994. **Colin Macilwain**

Japan: Recession threatens shift to basic research

Tokyo. While most countries around the world — including its immediate neighbours, South Korea and Taiwan — are promoting research aimed at developing new technologies, and cutting back on basic science, Japan, the pioneer of this strategy, is heading in the opposite direction.

But as the economic recession continues to bite and international competition becomes more fierce, Japan may find that it too comes under pressure to follow the worldwide trend in order to restore its economic growth. This will be a critical year for deciding which path it will follow in future.

Since the beginning of the 1980s, government and industry in Japan have both accepted that they should support basic research, reacting in part to criticism from the West, and particularly the United States, of the country's perceived tendency to live off the fruits of Western science.

In the latter half of the decade, some of this talk was translated into action. Companies flush with cash began setting up basic research laboratories at home and overseas, the latter often in close association with Western universities.

Similarly, government agencies such as the Ministry of International Trade and Industry (MITI) launched a number of projects aimed at supporting basic research. A typical example is the International Human Frontier Science Program (HFSP), established in 1989 by MITI and the Science and Technology Agency (STA) to support international research on the brain and biological functions, with the bulk of the funding coming from Japan.

Public pronouncements and actions by the government still support this push for basic research. Last year, for example, the Ministry of Education, Science and Culture continued its drive for large increases in the budget for university research grants — still very small compared with some Western countries — with the aim of almost doubling the budget in a few years.

Similarly in October, MITI, in collaboration with industry, began a large project on long-term research on nanotechnology. Several US semiconductor manufacturers have joined the project because, they claim, it is difficult in the current economic climate to win support for such research from their own companies or government.

In line with this trend, a number of government-funded research organizations have carried out reforms and introduced external reviews to strengthen their ability in basic research. First off the mark was Tokyo University, which, after persuading the government to increase funding for buildings and graduate research, brought in a team of ►

Italy draws order out of chaos

Munich. Last month, Italy's Ministry of Research and Universities announced plans to scrap the country's much-criticized system of distributing research funds through small committees, and to introduce for the first time a formal peer-review process for almost all government-funded research.

This is the latest move by the ministry to develop tighter control over the quality of the research that it funds and align its procedures for allocating research funds more closely with those of other Western nations.

The new system will be introduced gradually beginning in 1994, when three per cent of all research grants will be allocated after peer review. It will be followed by a scheme under which research projects receive inter-

national evaluation at defined intervals.

At the same time, discussions have begun between the ministry and Italian industry to identify long-term strategic research needs. The ministry has already decided to launch a strategic programme on robotics, once its budget for next year has been agreed. Also under consideration are programmes in environmental technologies, informatics and nanotechnology.

Much of the credit for bringing order into a reputedly chaotic ministry must go to the new minister of research, Umberto Colombo, who has been in office only since May. But he is not a politician, and therefore may be replaced after next spring's elections. If so, there are many in Italy who hope that the reforms he has started will be respected by his successor. **Alison Abbott**

foreign and Japanese scientists to assess its physics department in January 1993.

Similar reviews followed of the STA's Institute of Physical and Chemical Research (RIKEN), and the Ministry of Education's Institute of Space and Astronautical Science. And MITI has given its 15 institutes some autonomy in deciding how to spend their funds. In return it has asked them for more accountability, in particular the introduction of external reviews to ensure they remain at the forefront of basic research.

The situation in neighbouring Taiwan and South Korea is very different. Two years ago South Korea launched its "G7" project, under which the government and industry plan to invest 3,400 billion won (US\$4.2 billion) by 2001 to catch up with the technological achievements of the world's seven leading industrial nations.

Some South Korean researchers carrying out very basic research in fields such as structural biology have managed to win funds from this project. But its main focus is on specific technologies, such as high-definition television, megabit semiconductor chips, new materials and biotechnology.

Similarly, the Taiwanese government, which has traditionally supported university and academic research much better than its neighbour Japan, has introduced a policy of supporting research aimed at specific technological goals.

The National Science Council (NSC) has cut funds for investigator-initiated grants by 20 per cent, and is diverting the money to university-industry consortia for "mission-oriented" research in ten areas, ranging from consumer electronics to aeronautics. The NSC says it is doing this because university researchers have placed too much emphasis on publishing papers and pursuing theoretical studies.

Japan, which in the past has been in the forefront of countries pushing technology-oriented research, is under pressure to return to this approach. Government officials such as those who helped launch the HFSP are becoming concerned that, because of the tendency of their colleagues to use developments in the West to argue for new projects, there may be a swing back to more applied research just as efforts to support basic research are getting off the ground.

There are similar fears in industry. Many Japanese companies set up well-funded basic research institutes during the boom of the late 1980s. Their official line is that they have no intention of cutting back on basic research, and that any cuts in spending have been part of across-the-board cuts for research and development; if anything, they claim, basic research has suffered the least.

But researchers in these institutes fear that as the recession continues to bite management may begin to look on these new institutes as a luxury they cannot afford — or at least to direct their research more towards applications. **David Swinbanks**

Europe flexes its muscles

Paris. The joint research efforts of the 12 member states of the European Union (EU) jumped from relative obscurity into the political spotlight at the end of 1993. This year will see how much talk is translated into action.

At their summit meeting in December, the EU heads of states endorsed a white paper (policy document) from Jacques Delors, the president of the European Commission, giving considerable importance to support of research.

Two events in 1993 changed the face of EU research politics. The Maastricht Treaty came into force in November, giving new powers to the European Parliament. The main change is that the parliament, like the member states, can now veto the Framework budget; and it has been quick to insist that this power of veto carries the right to shape EU policy from the outset.

Reinforcing this message, the parliament organized a big European Science Summit in Brussels in October. The take-home message was that it wants the fifth Framework, which will start in 1998, to support more research into the social and economic impact of science. The parliament also wants more spent on research in renewable energy, life sciences, environment and new industrial technologies.

The other twist was the commission's decision to delegate management of an entire ECU24 million (US\$27 million)

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Unrivalled: with the SSC's demise, eyes are on Europe's Large Hadron Collider.

plant molecular project to a group of scientists. If the experiment succeeds, the commission will probably set up similar structures in other research areas. This signals an important shift in policy. Previously, suggestions that bodies such as the European Science Foundation could administer EU research programmes have been quickly dismissed by the commission.

Any development to decentralize research administration from Brussels would tie in neatly with Delors' plans to encourage EU member states to do more to coordinate their national science policies. **Declan Butler**

Germany: still feeling the pinch

Munich. Germany has long been one of the world's biggest spenders on research and development. Indeed, for several years it has come top of the league for spending on civilian research as a proportion of gross national product. But the picture is no longer so rosy. With reunification continuing to drain coffers already under pressure from the world recession, government research budgets have been frozen.

Last year, the Federal Ministry of Research and Technology (BMFT), which handles over half of the total government research budget, had to make do with the same amount of cash — around DM9.4 billion (US\$5.5 billion) — as in 1992. This figure will not increase significantly in 1994; and with inflation expected to be over 4 per cent, that effectively means a cut in spending.

The government wants industry to benefit much more from public investment in research. So far, and despite urging from certain sectors of industry and the country's Social Democrats, it has resisted pressure to achieve this aim merely by shifting funds from basic to applied research.

In contrast, the government's main con-

cern at present is to increase the efficiency with which research results are transferred out of the laboratory into industry. Funding of basic research, at nearly 40 per cent of the BMFT budget, is generous. But Germany is aware that it has been slower to turn the products of research into wealth-creating products than its main competitors, the United States and Japan.

In addition, Germany faces public hostility towards key technologies, particularly biotechnology and genetic engineering. Expressed through stringent legislation controlling genetic experiments, this hostility has been a major disincentive for industry to invest in this area.

Chancellor Helmut Kohl's coalition government is keen to address both issues, and to create a more friendly environment for both industry and applied research. The gene laws have already been revised to make them less restrictive. And Paul Krüger, the research minister, has various plans to bring researchers and industry into closer contact.

Last summer, for example, Krüger set up a technology council made up of twelve representatives from industry and the ►

scientific community, to advise him on how to develop a more coordinated and responsive research strategy. One recommendation of the council, which met for the first time in September, was that the research minister should have broader responsibilities.

In particular, the council has said it would like the minister to have more control over the total government research budget; at present, responsibility is divided between at least four ministries. It now has the task of suggesting how the desired closer contact between science and industry can become an effective reality.

The creation of the technology council has prompted lobbying from groups keen to see a change in policy. At the end of November, for example, the Christian Democrats published a 22-point proposal for a more strategic allocation of research funds.

This was followed in December by recommendations from the Bundesverband der Deutschen Industrie, German industry's umbrella organization, on topics the technology council should take up to improve strategic thinking about research. Most of the recommendations coincide with the council's own initial conclusions.

The growing interest in science and technology policy has also prompted a parliamentary debate due to take place next week (13 January). Krüger is expected to take the opportunity to insist that, despite previous rumours to the contrary (see *Nature* 365, 481; 1993), there will be no shift in the proportion of research ministry funds going to basic research in the near future. German scientists are approaching the new year cautiously, but perhaps with less foreboding than in the recent past. **Allison Abbott**

France shifts the focus from military to civilian technology

Paris. Shortly before Christmas, a report to the French government recommended that France give up its autonomy in armaments manufacturing, and share its costs with other European countries.

Such a proposal would have been unthinkable a few years ago; that it is now being seriously considered demonstrates the impact that changes in the world's geopolitical scene have had on French thinking about its technology strategy — and, indirectly, on its support for science.

France ended almost 12 years of unbroken Socialist rule in March 1993 by voting decisively for a conservative government. Over the next few weeks, the government will set about developing a national research strategy through a vast national consultation with a debate in parliament.

Although the new government is taking the credit for this initiative, much of the demand for change has arisen from policies launched by Hubert Curien, the minister of science in the previous Socialist government, and has been prompted by geopolitical adjustments to the end of the Cold War.

These adjustments have in particular removed the justification for much of the research that has assured France's military might, and earned it a disproportionately strong voice at the world negotiating table. As a result, France is having to reshape its research base into a tool for promoting industrial rather than military competitive-

ness, swapping its military goals for a desire to protect and exploit its strengths in such areas as computer software, biotechnology and the audiovisual sector.

Change in 1994 will also focus on longstanding structural problems in the research and university system and will be intended to stimulate the productivity of the science base. Current problems include poor coordination of national research policy, and lack of mobility of scientists both within the research organizations and between the universities and industry.

The national consultation is unlikely to bring much radical new thinking to these problems. But, like the ensuing parliamentary debate, it should help to crystallize potential solutions, and stimulate public awareness of the importance of research to the economy.

The most likely outcome will be an enhanced trend towards more strategic research, achieved not by simply increasing support for applied science, but by making basic research and technology-transfer more directed and efficient.

Basic science will receive some protection merely because almost all French researchers are civil servants on whom the government is reluctant to impose job reductions. Nonetheless, observers predict that areas of long-term basic research will probably stagnate.

On the ground, 1994 will see a big increase in the number of small mixed laboratories bringing together industry and government-funded scientists. The FFfr 1.6 billion (US\$270 million) government-backed biotechnology programme, BioAvenir, that links Rhone-Poulenc with a clutch of public laboratories, is likely to become a model for close cooperation between industry and public science (a similar programme for chemistry, ChemiAvenir, is already on the cards).

Intriguingly for a government pledged to loosen the state controls of its predecessor, the task of ensuring that industrial competitiveness becomes a priority for the whole research system will be achieved in 1994 by stronger central control of strategy. The Centre National de la Recherche Scientifique (CNRS), for example, is due to sign the first of the new five-year contracts under which the ministry and research organizations aim to agree on fixed objectives and the means to achieve them (see *Nature* 365, 598; 1993).

François Kourilsky, the director general of CNRS and a Socialist supporter, steps down in June. CNRS management will probably replace him by someone more sympathetic to the government's aims, and more radical reforms may follow.

Declan Butler

Russian election re-opens old conflicts

Moscow. Russian science, like much else in the country, enters the new year in a state of turmoil and uncertainty. Much of this results from last month's elections, one of whose immediate effects could be the virtual disappearance of the Ministry of Science, and a decision to return responsibility for basic science to the Russian Academy of Sciences.

During the past year, the ministry has been at the forefront of efforts to reform Russia's scientific system, bringing its practices more in line with those of the West by, for example, encouraging greater competition between researchers and closer links with industry.

Shortly after the elections, however, President Boris Yeltsin announced plans to make drastic cuts in the number of ministries. According to some sources, the Ministry of Science and Technical Policy leads the list.

Closing the ministry was first proposed to the president in September last year in a memorandum by Nikolai Malyshev, his

adviser on science and education. Malyshev proposed putting the academy back in charge of all basic research, and merging the Ministry of Science — which he described as functioning in the "old-fashioned" Soviet way — with the Ministry of Education and the State Committee on Higher Education to form a single Ministry of Science and Education.

Ministry officials and their supporters deny Malyshev's charges. They claim that, because of low salaries, the number of staff working in the ministry is actually decreasing. They also reject the criticism that newly-created bodies such as the Russian Foundation of Basic Research waste money in the classic Soviet way.

Academy officials argue that the organization needs to be given new authority to restore the fortunes of Russian science. Its critics claim that such a move would, in practice, do the opposite. The debate is guaranteed to intensify in 1994.

Vladimir Pokrovsky

Britain awaits impact of 1993 reforms

London. In few countries was the transition of science policy in 1993 marked as clearly as in Britain, where the publication in May of the government's white paper (policy document) on science and technology proved, as previously anticipated, to be a turning point in the country's approach to the organization of science.

The white paper was broadly welcomed, even by those who disagreed with some of its conclusions, as the first government-wide statement of science for almost 20 years. It was also seen as an explicit acknowledgement of the political importance of research in a competitive global economy.

But the document's impact lay less in its relatively straightforward message — that science deserves continuing support from public funds primarily as a source of long-term wealth, rather than as an unrestricted search for new knowledge — than in the procedures it proposed for carrying these out.

Three changes stand out. The first has been the reorganization of the research councils — in particular the splitting up of the former Science and Engineering Research Council — giving each new council a clear mission which makes explicit their responsibility towards the nation's economic health.

The second has been the decision to integrate the management of the research councils into the functioning of government by making them responsible to the holder of the new post of director general within the

Office of Science and Technology.

The third has been a broad-based exercise in crystallizing thinking about the country's technological objectives, known as "technology foresight" and designed to make the research community aware of where industry (and government) would like its future priorities to lie.

Government officials claimed at the end of the year that the political success of the white paper was demonstrated by its effectiveness in helping to ward off the Treasury's efforts to impose major cuts in public spending on research.

Industry, too, which is now being offered a greater role than ever in determining the country's research agenda, gave its general endorsement. Its main reservation was that it had not been sufficiently consulted over, for example, proposals for a four-year research doctorate or technology foresight priorities (both of which the government moved hastily to rectify).

Scientists were more cautious. With no promise of extra funding, and strong evidence of greater external control over how funds are to be spent, there was little concrete to delight scientists.

But it could have been worse. "What has come out is seen as moderate, potentially beneficial and certainly better than might have been expected" says Sir Michael Atiyah, president of the Royal Society.

1994 will tell how much of this potential

benefit will materialize. It will also show whether the changes being sought by the government are sufficiently widely accepted to survive future short-term shifts in national priorities.

In the immediate future, both the new research councils and the higher education funding councils face the task of devising an effective measure of the potential usefulness of strategic research (rather than merely scientific merit) that can be used to judge individual grant applications.

Many researchers are worried that excessive heavy handedness in applying cost-benefit analysis to do this, however sophisticated, will stifle long-term creativity. "As in physics, there is a danger that if you start observing a system too closely, you may interfere with its effects" says Atiyah.

A different challenge facing the government is how far its commitment to market forces should apply to institutions such as the National Physics Laboratory and the Laboratory of the Government Chemist, each founded on the philosophy that there are some national needs which the market is unable or unwilling to address.

"This is the dark side of the white paper," says Valerie Ellis, deputy general secretary of the Institute of Managers, Professionals and Specialists, which has expressed alarm at recent reports that many of these laboratories are going to be fully or partially privatized in the near future. **David Dickson**

Commercialization is the word in India

New Delhi. 1993 was a bad year for Indian science. Budget cuts, inflation and the devaluation of the rupee left research institutions with a reduction of 20 to 30 per cent in research funds. No new major projects were launched, and the country's largest research body, the Council of Scientific and Industrial Research, was only just able to pay salaries and electricity bills.

This year may be better. It is unlikely to see an increase in taxpayers' money going to publicly funded laboratories. But more money may come from industry. A new technology policy to be introduced early in the year requires each company to place a percentage of its annual turnover in a research and development fund.

In the public sector, projects will receive funding only if they yield marketable products, says S K Joshi, chief of CSIR. Two factors have led to this approach: liberalized economic policies, which have increased competition; and the revisions to the General Agreement on Trade and Tariffs, which will now prevent Indians from copying products developed by others. **K. S. Jayaraman**

Canada promised a new boost for R&D

Ottawa. Canada's science policy in 1993 was dominated by the change of government, and the promises made by the Liberal party before the election seem certain to set the policy directions for the new year.

Soon after the election, Jean Chrétien, the new prime minister, reaffirmed his intention to introduce a series of measures designed to boost research and technology in Canada, despite a budget deficit C\$5 billion (US\$3.8 billion) higher than the previous government's forecast.

The measures include doubling investment in research and development, with emphasis on small and medium-sized businesses, provided Canada shows it can manage the increase. Technology networks would be established between universities, industry and government and co-operation and partnerships encouraged between them.

Chrétien said that the new government will also continue to support basic research, to include "stable funding for the granting councils, the National Research Council, and the Networks of Centres of Excellence".

David Spurgeon

Australian scientists feel the heat

Sydney. The Australian government's long-standing policy to make all the money it spends on science outside the universities count towards economic growth was further consolidated during 1993.

One element of this strategy was increased pressure on the government's main research organization, the Commonwealth Scientific and Industrial Research Organization (CSIRO), to find more of its budget from the industries it was formed to serve. Thus its budget was cut by 1.3 per cent, and a special annual grant of \$A50 million has come under threat.

In contrast, Co-operative Research Centres (CRCs) received more money. These are joint industry-university research centres designed to improve the commercialization of academic inventions and to stimulate industry-orientated research. More than 50 are planned.

For those scientists uninterested in commercial research there was some good news; the Australian Research Council received a 7 per cent increase in funding, to A\$294 million. **Mark Lawson**

Follicles in competition

SIR — Competition for the privilege of ovulation there certainly is, but to accept as chief candidates, as Daedalus does, the left and right ovaries, "locked in endocrine poker" (*Nature* 362, 502; 1993), is imaginative to a serious fault. Biological objects important to Jim Watson may have come in pairs, but the pairedness of a woman's ovaries demands no more mystical an explanation than anatomical symmetry plus a dab of insurance (should one twist on its pedicle and fail). Many fish, birds and bats do perfectly well with one¹.

The participants in the competition to ovulate each month are, in any case, much more than two. They consist of all those oocyte-bearing follicles, distributed across both ovaries, with between about 75,000 and 375,000 granulosa cells by the end of the previous luteal phase². What Daedalus divines as each ovary "raising its level of hormonal bidding, until one of them feels outbid, and folds", need be no more than stochastic stumbling between two ovaries which, for virtually all physiological purposes, act as one.

Competition among the follicles during a woman's life is, on the other hand, tough. Before the first ovulation, the oocyte population has fallen from 2 million at birth to 300,000 (ref. 3). If every cycle is ovulatory from then to the menopause, when almost no follicles are left, just 500 to 700 follicles will qualify each month. Each follicle hangs on central control by the hypothalamus and pituitary gland through the pituitary hormone, follicle-stimulating hormone (FSH). This, not the other ovary, is the command centre that needs to be convinced that there is a dominant follicle extant. Once that is done, and the dominant follicle has sequestered enough FSH to keep growing with much less new FSH than its neighbours, nothing more mysterious is needed to override the control than a constant dose of administered FSH. That is what we as specialists in infertility do every day of the week to induce multiple follicular development for helping conception with *in vitro* fertilization.

In natural ovarian cycles, FSH levels rise at menstruation to encourage growth of a new cohort of follicles (numbering perhaps 50 in the early 20s, maybe five at 40, probably just one at 48); then, as follicles respond and grow further, FSH levels fall — depriving more and more follicles of the FSH needed to stay in the race. The follicular signals that DREADCO's endocrinologists will be looking for are those that diminish FSH at just the right rate to leave just one victorious follicle (but often two follicles, sometimes three) still growing. The contenders for Daedalus's 'Anova', the putative endocrine bid for supremacy, are directed not

to the other ovary but centrally, and include the steroid hormone oestradiol, the well-known peptide inhibin⁴, and a substance called gonadotrophin surge attenuating factor⁵, known only to be not a steroid and to have a molecular weight between 10,000 and 30,000.

On the other hand Daedalus's intended infertility treatment 'Binova' — his "biochemical misère bid (to) convince each ovary that the other has folded (so that) both will therefore produce an ovum that month (and) double the subject's chance of pregnancy" — has been known for years: the oestradiol-antagonist clomiphene⁶. More effectively, though, we override the subtleties of inhibitory control with, simply, injections of FSH.

Robert Jansen

Sydney IVF,
187 Macquarie Street,
Sydney NSW 2000,
Australia

1. Van Tienhoven, A. *Reproductive Physiology of the Vertebrates* 119 (Cornell University Press, Ithaca, NY, 1993).
2. Gougeon, A. *Ann. Biol. Anim. Biochem. Biophys.* **19**, 1461–1468 (1979).
3. Bakert, T. G. in *Reproductive Biology* (eds Balin, H. & Glasser, S.) 398–437 (Excerpta Medica, Amsterdam, 1972).
4. Findlay, J. K. *Biol. Reprod.* **48**, 15–23 (1993).
5. Fowler, P. A., Messinis, I. E. & Templeton, A. A. *J. Reprod. Fertil.* **90**, 587–594 (1990).
6. Jacobson, A., Marshall, J. R., Ross, G. T. & Cargille, C. M. *Am. J. Obstet. Gynecol.* **102**, 284–290 (1968).

Why Africa said no

SIR — The African Academy of Sciences, in its dissent from a resolution on population growth passed by all but three of the 15 academies from the developing world assembled in New Delhi, showed more enthusiasm for politics than for science (see *Nature* 366, 3; 1993).

The World Bank's 1993 World Development Report shows the total of 24 countries of Africa south of the Sahara to have an overall 46 per cent of population under 15 with a total fertility rate of 6.4. Their high population growth rates have resulted in declining food production per caput for the past 20 years. Their best efforts to feed their peoples and to provide them with schools and with hospitals have been continuously overwhelmed by population growth.

The African Academy raised the question of "the contribution of the North to Africa's population predicament". This contribution has, for a century, been the very effective assistance in the reduction of death rates through medical services, clean water supply and agricultural progress. Unfortunately this help has been accompanied, until recently, by a confused reluctance to assist in the reduction of birth-rates. These African governments have in recent years shown awareness of

the problem and have invited cooperation from the United Nations, bilateral and nongovernment organizations in the development of family planning services. Thus although a small group of academic officials have made a political gesture, "Africa" has definitely not said "no" to a reduction in fertility rates.

Charles Pereira

Peartrees,
Nestor Court,
Teston,
Maidstone,
Kent ME18 5AD, UK

CNRS defended

SIR — As directors of research of the French Centre National de La Recherche Scientifique (CNRS) we regret the negative impression you give of this research organization (*Nature* 365, 95; 1993). Any national body of this kind comes in for criticism from a few individuals who do not agree with the assessment of their work by scientific committees, or who object to the relocation of their laboratories for practical reasons. For the CNRS, with more than 11,000 researchers, who are in the privileged position of having permanent contracts, it is essential that the processes of scientific review and on occasion consequent closure of laboratories and reorientation of personnel should be carried out correctly. Scientists and laboratories are reviewed every two years by national committees whose members are nominated by the CNRS and elected by the scientific community.

In our experience, this process functions properly. We do not know of any scientist who has been dismissed for publicly criticizing government policy or the administration of research. Scientists can be dismissed for proven professional misconduct, but this is very rare. A lot of effort on the part of the scientific administrators is put into trying to revitalize flagging researchers. In the past the CNRS has been criticized for being too easy-going; with budgetary restrictions and a lot of good young scientists applying for jobs, the review committees have become "tougher". This can only be beneficial, even if it is resented by some sections of the community.

Margaret Buckingham (Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France), **Marcel Doree** (Montpellier), **Daniel Louvard** (Institut Pasteur, Paris), **Jacques Mallet** (Gif-sur-Yvette), **Jacques Pradel** (Université Aix-Marseille), **Alain Prochiantz** (Ecole Normale Supérieure, Paris), **Jacques Samarut** (Ecole Normale Supérieure, Lyon), **Mathias Springer** (Paris), **Patrick Stragier** (Paris), **Mary Weiss** (Institut Pasteur, Paris), **Jean Weissbach** (Généthon), **Moshe Yaniv** (Institut Pasteur, Paris)

Eighteen ninety-four and all that

J. L. Heilbron and W. F. Bynum

Understanding the intricacies of colour vision, the origins of the DNA industry and the classification of stars are but three of the scientific birthdays celebrated in this year's anniversarial feast.

ON 8 September 1894, German science suffered a terrible loss. At one stroke it lost its leading physiologist, its leading physicist and its leading spokesman. He was Hermann von Helmholtz, formerly a professor at the universities of Königsberg, Bonn, Heidelberg and Berlin, and lastly president of the Physikalisch-Technische Reichsanstalt, the German standards laboratory, which he had helped to found. He had contributed to medical practice (he invented the ophthalmoscope), to understanding of hearing and vision, to the mathematical theories of hydrodynamics, electricity and magnetism, to the popularization of science and its philosophy, and, through his friendship with Lord Kelvin, to the internationalization of scientific relations. We call particular attention to his theory of colour vision, based on that of Thomas Young, in turn based on the self-analysis of John Dalton, who called pink geraniums blue, as an advance especially worthy of celebration in 1994.

Helmholtz would have been the obvious choice for our award of 'anniversarian of the year' were it not for the hasty handlers of the guillotine who, in 1794, separated the head from the body of Antoine Laurent Lavoisier. Lavoisier was not only France's leading chemist, but probably its most important experimental physicist and its most influential technical expert. Alas! He was also a conspicuous and wealthy tax farmer — a partner with the government in a corporation to exact money from a resentful population. The several Lavoisiers would appear to have as strong a claim to be anniversarian of the year as the several Helmholtzes.

To adjudicate their claims, we have weighed every relevant scruple with the expertise gained by 7 years' exercise of the anniversarial art; and we have found the claims to balance more exactly than the products of chemical combinations in the experiments of Lavoisier. To tip the scales, we added to the multiple personae of each claimant that (or those) of his students who died the year he did. The principle delivers Helmholtz as anniversarian of the year, since, on new year's day 1894, his student Heinrich Hertz, who first showed experimentally the existence of radio waves, died from a painful cancer.

We proceed to a list of anniversaries eligible for celebration in 1994. As usual, we begin with centenaries, work backwards at 50-year intervals, and end with

more recent things. To save space, "c" will signify a century.

1894 (1.0 centenary)

Hertz and Helmholtz were not the only heavy blows to German physics in 1894. Almost midway between their deaths, August Adolph Kundt also died. Kundt was best known for his measurements of the speed of sound in gases in a squeaky glass vessel strewn with lycopodium seeds — the 'Kundt's tube' that has abused the ears of generations of physics students.

The losses among Germany's scientific rivals were less concentrated but nonetheless plentiful. France lost Ferdinand Marie de Lesseps, builder of the Suez Canal and would-be builder of the Panama Canal; the Swiss, Jean Charles Galissard de Marignac, a painstaking determiner of atomic weights and analyst of the rare earths, two of which, ytterbium and gadolinium, he more or less isolated; and the United States, Oliver Wendell Holmes, poet, wit and anatomist whose contribution to safer childbirth was noted last year. Old men everywhere lost Charles-Edouard Brown-Séquard, whose self-rejuvenation experiments with guinea-pig testicle extracts did not yield their anticipated results.

Further to the anniversarian of the year and the runner-up, Arthur König, a student of Helmholtz's, who had been working on his teacher's theory of colour vision since 1884 (1.1c), discovered the role of visual purple in sight; J. von Kries advanced the theory that the rods and cones of the retina constitute two separate visual

systems; Hans Landolf confirmed what Lavoisier taught, that weight is conserved in chemical reactions; and Max Rubner (born 1854, 1.4c), confirming Helmholtz as well as Lavoisier, demonstrated definitively the applicability of the law of conservation of energy to living organisms.

Earth scientists have much to celebrate this year. They might mark the centennial of Albrecht Penck's *Morphologie der Erdoberfläche*, the first unified text of geomorphology. Rising higher, they might notice that the Deutsche Verein zur Förderung der Luftschiffahrt sent an unmanned balloon to 18,450 metres and a temperature of -67°C . Dropping lower, they might make merry over the publication of Franz Krauss's *Höhlenkunde*, which opened a new era in the scientific exploration of caves.

Again in the air, Angelo Mosso, using soldiers as guinea pigs, issued a revised edition in German of his book of 1884 (1.1c) on the effects of altitude on Italian army units on Monte Rosa. Again on the ground, the British dug up lots of it to finish the Manchester Ship Canal, which, had it been larger, would have allowed the newly launched *Turbinia*, the first ship propelled by a Parson turbine, to reach Manchester from the sea; and Marie Eugène Dubois removed soil in Java to reveal *Pithecanthropus erectus*.

The importance of the creation of Henry Wellcome's Physiological Research Laboratories in London, and of the Institut Pasteur in Tunis, which Charles Nicolle was to direct so imaginatively for 34 years, should not be missed; nor the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Left, Helmholtz, anniversarian of the year for his talents and protégés; right, Samuel Morse.

Hulton Deutsch/Mary Evans

potential value of the simultaneous discovery of the plague bacillus by Shibasaburo Kitasato, who had studied with Robert Koch, and by Alexandre Yersin, who had taken up bacteriology after Emile Roux saved his life with rabies serum; nor, perhaps, the promise of the cinematograph, invented by Ernst Lumière. More mundane was Raoul Pictet's finding that snails can withstand temperatures of -120°C .

1844 (1.5 centenary)

"What hath God wrought?" The famous message transmitted between Washington and Baltimore by Samuel F. B. Morse, a painter turned inventor, on 24 May 1844 demonstrated the practicability of land telegraphy over useful distances. God also wrought that year, through his servant Charles Goodyear, a prototypical inventor who worked in poverty and realized little from his genius, the first pair of vulcanized rubber shoes; through John Mercer, a calico printer in Lancashire, the process of mercerizing cotton; through a Scottish publisher who preferred anonymity, *Vestiges of the Natural History of Creation*; and, through Julius Bodo Unger, the principle of guano.

Towards the end of 1844, tests began on the gigantic reflecting telescope on which William Parsons, third Earl of Rosse, had worked for 20 years. The mirror, whose casting from an alloy of tin and copper was Rosse's main contribution, weighed 4 tons and had a focal length of 54 feet; the mounting and drive machinery, which weighed more than 150 tons, were built on Rosse's estate by his peasants, who thereby learned not only mechanical skills but also, we are assured, "steady habits and good conduct". The instrument served Rosse better than the two refractors with which Percival Lowell stocked his 1894 (1.0¢) observatory in Arizona: for Rosse resolved nebulae into their individual stars and discovered the spiral character of some of them, whereas Lowell detected traces of civilization on Mars.

Many will want to remember John Dalton, who died 150 years ago. His centre of operations was the Manchester Literary and Philosophical Society, of which he became a member in 1794 (2.0¢). In most years Dalton would be celebrated primarily for his atomic and molecular theories. But in 1994, we recommend instead an emphasis on daltonism (see 1794). Physicists who prefer comings to goings should know that as Dalton departed this life Ludwig Boltzmann entered it.

1794 (2.0 centenary)

In a fit of revolutionary inconsistency, France's new republican government charged the Académie des Sciences of Paris with completing the metric reform in 1794 and abolished it (the academy, not the metre) 2 months later. Then came the

Theiwell and Lymm cutting of the Manchester Ship Canal, completed in 1894.

abolition of academicians like Lavoisier compromised by prominence in the *ancien régime*. Lavoisier's young widow, whose drawings constitute a rare depiction of laboratory life in the late eighteenth century, lived to suffer further in the cause of science by marrying the American tory Benjamin Thompson (died 1814, 1.8¢), established Count Rumford of Bavaria in 1793 (2.0¢ to 0.5%), whose experiments on the grinding of gun barrels will be conspicuous in these pages in 1997.

Dalton's first paper before the Manchester Literary and Philosophical Society reported his inability to distinguish red from green; an inability that, according to legend, he discovered when his coreligionists at a Quaker meeting objected to his bright red stockings, which he thought a sober grey. He ascribed his problem to a blue tint to his eye. His countryman, Thomas Young, who also provided the clue to unscrambling Egyptian hieroglyphics, instead attributed it to the want of red detectors in the retina which, on his theory, normally contained red, green and violet receptors. It was this theory that Helmholtz revived and refined in 1852, 50 years (0.5¢) after Young proposed it. Dalton would have nothing of Young's notion, and directed that his eye be dissected after his death — significantly, in 1844 (1.5¢) — to demonstrate the azure tints to which he ascribed his handicap. Nothing of interest was found.

In two other throw-forwards to eligible anniversaries, Johan Gadolin, the spokesman for Lavoisier's chemistry in Scandinavia, completed in 1794 his analysis of a black rock from the village of Ytterby in Sweden, from which Marignac (see under 1894) extracted two earths, one of which he called gadolinium, the first element named after a person. And Ernst-Friedrich Chladni interrupted his studies of the vibrations of plates, which he demonstrated using sand in a manner similar to the technique Kundt would adopt (see again under 1894), long enough

to argue persuasively his novel view that meteorites come from the sky.

Mention of the academy reminds us that, in 1794, its *secrétaire perpétuel*, the Marquis de Condorcet, failed the test of perpetuity. A mathematician by academic profession, he tried his hand at politics during the revolution. His draft constitution for the Republic was rejected; and he criticized in print the version that superseded his. An order was issued for his arrest; he fled, and in hiding wrote an extraordinarily upbeat book, *Esquisse d'un Tableau des Progrès de l'Esprit Humain*. Afraid of being discovered, he left his sanctuary and was caught; he died in his cell on the night of his arrest.

We must not omit to record, under more creative work of the new Republic, the foundation of the élite engineering school the Ecole Polytechnique (whose graduates, it is said, know everything and nothing else); and to note the birth of the polymathic William Whewell, whose foible was omniscience.

1744 (2.5 centenary)

Pierre Louis Moreau de Maupertuis' natural moroseness was deepened by tramping the high Andes to find the size of the Earth and by suffering the lampooning of Voltaire. Nonetheless, in 1744 he performed two significant acts of irenicism. For one, he joined Frederick the Great's Academy of Sciences in Berlin, thus placing French science at the pinnacle of the German academic enterprise. For another, he declared the principle of least action, as a synthesis of Fermat's principle of least time and Descartes' mechanical theories which all *cognoscenti* had thought irreconcilable. For irenicism, however, this could not compete with the magnificent gesture of the Roman Catholic Church, which this year permitted the reprinting of Galileo's *Dialogo* (1632), though it remained upon the index of prohibited books, provided that it thenceforth contain Galileo's recantation of its message.

About the time that Maupertuis was describing least action to the Berlin Academy, Mikhail Vassilievich Lomonosov was, perforce, observing it, in the prison to which his colleagues had referred him for his hot-headed style of argument. The imprisonment began in 1743; with this preparation he became, in 1745, the academy's professor of chemistry; in between he developed an atomic theory of matter inspired by Boyle's law and reached the doctrine that heat is a mode of motion. If Lomonosov's kinetic theory is to be celebrated at all in our time, 1994 is the year to do it.

It is also a year to celebrate the fecundity of water. Abraham Trembly (died 1784, 2.1e) of Geneva enriched his impressions of Dutch life by studying ditch water. He thus discovered freshwater polyps, whose ability to regrow whole organisms from bits of cut-up ones amused and instructed the age. Trembly collected his results in 1744.

While Trembly examined ditch life, Bishop Berkeley issued a treatise on the medical virtues of tar water. This elixir supported a discussion of contemporary chemical theory, an attack on newtonian physics, declaration of the true nature of causation, and a bridge, via fire and light, to the spiritual interpretation of all phenomena. Like Berkeley's other writings on science, this treatise, entitled *Siris*, taught that scientific theories are at best useful descriptive hypotheses and cannot attain to the essence of things. His point of view deserves note at a time when the doctrine of the Big Bang has blown up some physicists into theologians.

This year 250 years ago Anders Celsius, the light of Swedish astronomy and physics, was extinguished at the age of 43; John Theophilus Desaguliers, a Huguenot refugee who became Newton's paladin at

the Royal Society of London died, aged 51; and Jean Baptiste Pierre Antoine de Monet de Lamarck, whose name was as long as the neck of the giraffe, was born.

1694 (3.0 centenary)

Although no doubt the founding of the University of Halle, and the death of the Bolognese physician and microscopist Marcello Malpighi, will inspire tercentennial observations, as they should, we recommend for unusual excitement the work of Rudolf Jacob Camerarius.

About the time that he became director of the botanical garden at the University of Tübingen in 1688, Camerarius took up the vexed question of the reproduction of plants. By isolating female plants from others and cutting off inflorescences in monoecious plants, he showed that the anthers represent the male and the ovary the female parts of the flowers. In 1694 he screwed up his courage to report his findings in a letter, *Epistola . . . de Sexu Plantarum*. Despite the evidence, some botanists still refused to recognize the sexuality of plants when Queen Victoria ascended the throne. The matter was put to bed by the 9,000 hybridizations carried out by Friedrich von Gaertner, who reported his main results in 1844 (1.5e).

1644 (3.5 centenary)

Although no one has believed more than a word of it for two centuries, Descartes' *Principia Philosophiae* (1644) was one of the most important scientific books ever published. It first offered a radical and full alternative to the system of Aristotle. Where Aristotle overfilled the world with irreducible active essences of things, Descartes needed only inert matter in motion. For a century, European savants thought that the best guide to the operations of nature was the collection of *tourbillons* or vortices, which, according to Descartes, carried the planets, caused gravity, pushed the tides and brought about magnetism, electricity and the motion of animals.

Another book of great importance in science, Evangelista Torricelli's *Opera Geometrica*, also appeared in 1644. In it he took up topics of interest to his teacher, Galileo (free fall and projectile motion), and to his teacher's inspiration, Archimedes (the volumes of solid and properties of the parabola). In these investigations, Torricelli made good use of Cavalieri's method of indivisibles, an important forerunner of the infinitesimal calculus. Some of Torricelli's results in this line were transmitted through Isaac Barrow to his student Isaac Newton, who knew how to appreciate them.

We note the death of William Gascoigne, an inventor of the cross-hair micrometer for telescopic lenses, a great loss to English astronomy at the age of 32; and that of Johannes Baptista van Hel-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Charles Nicolle, first director of the Institut Pasteur at Tunis.

mont, a Flemish aristocrat, a follower of Paracelsus and the identifier of gas as a distinct state of matter, at the age of 65.

1594 (4.0 centenary)

In 1594 there passed from this terraqueous globe the best known of all geographers, he of the odd projection, Gerardus Mercator. Artist and cosmographer, philosopher and theologian, Mercator drew a great many maps, in fine style and unusual accuracy, which his son published the year after his death as an *Atlas*, the first use of the word to designate a collection of maps. Mercator introduced his projection in 1569. It made possible the easy determination of the rhumb, or constant compass bearing, on which a navigator had to sail to reach a distant destination without changing course. The convenience of obtaining bearings by laying a straight edge on a plane chart was bought at the price of large distortions of land masses to the extreme north and south.

1544 (4.5 centenary)

While on geography, we should recall that in 1544 Georg Hartmann (died 1564, 4.3e) wrote Duke Albert of Prussia a letter describing his discovery, made years earlier in Italy, of the dip of compass needles in the Earth's magnetic field. He also then made public his estimate of the declination of the needle at Rome, the first such determination on land. Another important event in the history of geomagnetism occurred in 1544, the birth of William Gilbert, whose *De Magnete* (1600) showed with the help of a spherical loadstone how the dip and declination arose and, into the bargain, offered an entire natural philosophy based on the Earth's magnetic (and electric) "virtues".

We are not done with geography. About 1540 Sebastian Münster, with a reputation in Hebrew, switched his studies to geography. Four years later, the first edition of his *Cosmographia* was "a description of the whole world and everything in it".

The Lord works in strange ways. Having informed his congregations that he had

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Lord Rosse's great reflecting telescope.

computed on irrefragable cabalistic principles that the world would end on 18 October 1533, Michael Stifel found himself out of job. He was rescued by his friend Martin Luther, who got him a new pulpit and an old wife. Thus cured of prophesying, he turned his mathematical abilities to algebra, into which he introduced the concept of exponent, a rudimentary sort of logarithm, and an effective symbolism, and published it in his *Arithmetica integra* in 1544.

The establishment of botanical gardens at universities was a development of almost inestimable importance because they, together with libraries, helped to push universities into owning and maintaining real property. One would like therefore to know when the first such garden came into being. We have unearthed two contenders, the University of Pisa, which claims that its garden dates from 1543, and that of Padua, which offers evidence of a garden in 1545. We suggest 1544 as the definitive average date of the first university botanical garden.

1944 (0.5 centenary)

The world being still at war, its major inventions served the purpose of annihilation: the Clinton uranium pile started up to produce plutonium for an atomic bomb; and the V-1 and V-2 rockets fell on Britain in a vain attempt to terrorize its citizens. Towards saving rather than destroying, the Johns Hopkins surgeon Alfred Blalock (died 1964, 0.3€) performed on 29 November 1944 the first 'blue baby' operation on a 2-year-old child suffering from Fallot's tetralogy.

A happier consequence of wartime circumstances was Wilhelm Heinrich Walter Baade's classification of stars into type I and II. As an enemy alien working at Mount Wilson in California, Baade could not perform the war work that occupied other astronomers; hence he had little competition for time on the telescopes which, because Los Angeles dimmed its lights at night, were more than ordinarily effective. In a timely reminder that *tempus fugit*, the Royal Observatory at Greenwich installed the first quartz-crystal clock to divide time ten times more exactly than the best pendulum mechanism.

Half-a-century ago, a physics laureate, Erwin Schrödinger, puzzled over *What is Life?* In the same year, a biologist published one of the most important modern papers not awarded the prize. Oswald T. Avery and colleagues identified deoxyribonucleic acid as the chemical that converted non-virulent strains of *Pneumococcus bacilli* into virulent ones and showed that DNA was the factor of this inherited variation. Avery lived to read Watson and Crick's paper of 1953, though not to see the first of many Nobel prizes the DNA industry has claimed.

Death took Arthur Stanley Eddington,

a British astrophysicist who championed the general theory of relativity (see under 1944±25); Charles Glover Barkla, the English discoverer of characteristic X-rays; Alexis Carrell, the French-born US pioneer of vessel surgery and propagator of tissue cultures; and Leo Baekeland, the Belgian-born US inventor of Bakelite. There came into the world paper

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

First or second? Plan of the garden at Padua.

chromatography and Oskar Morgenstern and John von Neumann's *Theory of Games and Economic Behaviour*.

1944 ± 25 (0.75€ and 0.25€)

Energies, pent-up or otherwise, directed during World War I were released to the advantage of science in 1919. Ernest Rutherford, returning to his laboratory, shot alpha-particles into the air and detected the first nuclear reaction intentionally induced by humans; Francis Aston made the first effective mass spectrograph; Vilhelm Bjerknes applied his new theory of weather fronts (inspired by the spectacle of trench warfare) to the formation of cyclones; Arnold Sommerfeld published the first edition of what immediately became the bible of the old quantum theory, *Atombau und Spektrallinien*; Bertrand Russell brought out the *Introduction to the Philosophy of Mathematics* that he had written while in prison for the sin of pacifism; British expeditions organized chiefly by Eddington to observe a total solar eclipse brought back results that appeared to confirm the predictions of another pacifist, Albert Einstein, on the basis of his general theory of relativity; and Robert H. Goddard published his *Method of Reaching Extreme Altitudes*, an inspiration to inventors of the V-1 and V-2 bombs (see 1944).

Many of the builders of later nineteenth-century science died. Britain lost William Crookes, chemist and spiritualist, who may thus have found his eschatological speculations confirmed;

and John William Strutt, third baron Rayleigh, who, among much else, almost killed the study of acoustics with his definitive *Theory of Sound*. The United States lost Wallace Sabine, whose design of concert halls showed that Rayleigh had left something for others to do; and Edward Charles Pickering, astronomer and astrophysicist. Germany had to do without Emil Fischer, chemistry's second Nobel laureate, for his contributions to organic synthesis and the German chemical industry, and Ernst Haeckel, Darwin's most ardent European disciple. Hungary mourned Roland Eötvös, who had showed definitively the equivalence of inertial and gravitational masses before Einstein took up relativity; Sweden lamented Johannes Robert Rydberg (born 1854, 1.4€), whose systematization of spectral regularities provided a prime ingredient for the development of the quantum theory of the atom; and Switzerland lost Alfred Werner, inventor of coordination chemistry.

The Nobel prizes for physical science presented more excitement than usual in 1919. None had been awarded in 1918. The reserved physics prize for that year went to Max Planck, for "his discovery of energy quanta"; the reserved chemistry prize, to Fritz Haber, "for the synthesis of ammonia from its elements"; the physics prize for 1919, to Johannes Stark, "for his discovery of the Doppler effect in canal rays and the splitting of spectral lines in electric fields". An ill-assorted group represented Germany in Stockholm in 1919: the patriotic Jew Haber, who had damaged his reputation by helping to implement the use of poison gas during the war, and who would be driven from the fatherland by the Nazis in 1933; the moderate lutheran Max Planck, who would continue to represent the best of German science under the Nazis; and Johannes Stark, who became a Nazi and the implacable foe of his fellow laureates.

Much worthy of celebration happened in 1969. But none of it can compete with that small step for a man, but gigantic leap for mankind, which occurred when Neil Armstrong put his foot on the moon.

1869 (1.25 centenary)

Rather than waste this space, we use it to draw attention to an anniversary of a kind we do not generally notice, one half way between a centenary and a sesquicentennial: the 125th anniversary of the foundation of a journal that ever since has been a premier repository of science worth celebrating, a journal that takes all of *Nature*, and something of history, for its subject. □

J. L. Heilbron is vice chancellor of the University of California, Berkeley, California 94720, USA; W. F. Bynum is at the Wellcome Institute for the History of Medicine, London NW1 2BE, UK.

What birthdays should be celebrated?

***Nature* is planning to celebrate its 125th anniversary this year with a series of events that may reinforce its international and public reputation.**

NATURE will be 125 years old on 4 November this year and plans to mark the occasion in several ways but with appropriate diffidence. As our regular (and self-styled) anniversarists imply in the very last paragraph of their article on page 11, a 125th anniversary (which so obviously almost escapes their attention) is not calculated to excite the imagination generally, as do anniversaries whose ordinal numbers are divisible by one hundred. But that is too narrowly based a proposition. In earlier times, the magic number might just as well have been 7, 13 or any other integer, suggesting that *Nature's* way out of this dilemma would be to invite its friends to a decent dinner on 4 November each year and leave it at that.

There are nevertheless good historical reasons for paying attention to this year's birthday, all of them to do with the present condition of science and with general expectations of it. A century ago, *Nature* had won a reputation as the champion of Darwinism and of rationality in the explanation of natural phenomena, but had not become a scientific journal in the modern sense. (Even the discovery of the electron still lay ahead.) Instead, the first editor's helpers, people like T. H. Huxley and John Tyndall, wrote their hearts out in telling what excited them in science. Presciently, as it turns out, *Nature* then as regularly berated the British government for its neglect of science as it does now.

But *Nature* is no longer a British journal. How could it be when almost 90 per cent of its readers are elsewhere? Indeed, from the start, *Nature* took the view that science is international, regaling its readers with the agendas of the academy meetings at St Petersburg and Philadelphia as regularly as with those of the Royal Society (of London) and of the Académie des Sciences (of Paris). *Nature* is now printed once a week at four centres (in Britain, the United States, Tokyo and Beijing) and once a month (as *Monthly Nature*) in Moscow.

The events of the very recent past can only reinforce that cosmopolitan tendency. All of us know of bright people who have been driven from Russia by repression (or recently, uncertainty) and from India's Indian Institutes of Technology by deprivation, and who are now card-carrying honoured members of the international community. Could even a low-grade 125th anniversary celebration help to deepen these international connections? (*Nature's* fondest wish is that there were another name for

"English" — "Esperanto" is bespoken — but that it were permissible to write it with a British accent.)

There is another pressure towards the celebration of this anniversary — the gulf that has recently emerged between science and those whom scientists believe should be science's beneficiaries, people at large. The past few decades have, it is true, seen a powerful growth of the health and wealth that people enjoy, much of which is a consequence of science and its many applications. But recent decades have also seen a growth of suspicion of science, especially in the rich countries of the world. (People in India or China, for example, see things very differently.)

Over the years there have been several valiant attempts to change this state of affairs. It is agreed that, in the long run, the solution lies in a more general understanding of the roots of science and also of its role in the remarkable history of this century, now almost past. But that may be too distant a goal. Should not a journal such as this, which benefits so much from its close relationship with the research community, be more directly engaged in helping to give currency to what is now exciting?

These two principles, the international character of science and its role in the general development of society, guide the plans so far devised for celebrating this year's birthday. Mostly the intention is to mark the occasion with a number of events, most of them in the second half of 1994. The centrepiece will be a conference on the general theme of how our world has evolved, to be held in New York on 31 October and in London on 3 November. We shall of course be as much concerned with physical as with biological evolution; it is hoped that the same people will speak on both occasions. With luck, these events should contribute to the general appreciation of our place in nature.

It is also planned to hold two one-day symposia on the mainland of Europe. The first, in Berlin, will deal with contemporary problems in genetics. The venue has been chosen because of the difficulties encountered in Germany in recent decades with matters such as legislation on genetic manipulation. Can informed discussion help to resolve these, or are they a foretaste of things to come elsewhere?

We also plan a symposium in Paris, later in the year, on the theme of the distinctiveness of science in Europe (where, after all, modern science began). That venue

marks the recent resurgence of science in Europe and the part played in that encouraging development by the consistency of French public policies on science over recent decades. Can the rest of Europe follow suit?

We have not forgotten Central Europe and Japan, where there are also tentative schemes that will introduce to audiences there people from elsewhere who may have interesting things to say about contemporary problems. These plans depend on the outcome of discussions with institutions in the regions concerned; details will be published later.

The ambition to make a more direct contribution to the general understanding is best realized by the means that *Nature* best understands — publication. It is hoped that 1994 will see the realization of a long-cherished project to produce a collection of the miscellaneous contents of the issues of *Nature* over the past 125 years which are a remarkably rich record of the development of modern science.

But that is necessarily a domestic preoccupation. During 1994, *Nature* will also attempt more deliberately than in the past to foster the spread of intelligence about science and its implications, not so much as a publisher in its own right but in collaboration with others. One objective is to demonstrate that the research community can, by its own efforts, help to make its own intellectual birthright more generally appreciated in the world at large. There are, after all, several routes to understanding.

Explaining the intricacies of, say, the structure of DNA to the world at large may be less effective a reinforcement of the general appreciation than, say, helping to ensure that the significance of DNA is understood within the general culture. But those who battle in the front line for understanding, in the world's classrooms, need more assistance than they are given now. That is another field in which the research community, and *Nature* in particular, could do more to help.

Further details of these events will be published in the next few weeks, as and when they are decided. Meanwhile, *Nature* would welcome suggestions from readers who may have other schemes for celebrating an off-season birthday. The guiding principle should be modesty. After all, *Nature* does not wish to take the wind out of the full-throated celebration there will be 25 years from now.

John Maddox

Another December revolution?

Paul M. Grant

IN December 1986 I boarded a plane in San José, California, bound for Zürich, determined to unearth the facts behind the reports and rumoured confirmations of superconductivity at 30 K which had recently been published by Georg Bednorz and Alex Müller¹.

Eighteen hours later, I stumbled jet-lagged into the IBM Ruschlikon laboratory. There I found an agitated and preoccupied young man — Bednorz had just received the preprint of Kitazawa and Tanaka's confirmation², and, justifiably, was concerned that the rest of the world would run away with his discovery (although Bednorz and colleagues had submitted their magnetization results for publication a week before the Japanese in October, the paper³ did not appear until February 1987).

Within half an hour, my travel weariness had disappeared. The amount and quality of the data my Swiss colleagues had accumulated was far beyond that generally known at the time, leaving no doubt that superconductors with T_c as high as 40 K really existed. I returned home a true believer, with the peace of my holiday delightfully shattered, ready to spread the good news to my colleagues in IBM Almaden.

I relate this story because in the middle of another December — last month — we again heard news of incredible advances in superconductivity. Within literally days of each other, two French laboratories reported evidence^{4,5} suggesting that superconductivity may occur above 250 K, perhaps as high as 8 °C. But unlike in the 'December revolution' of 1986, it is not at all clear yet whether the barricades of superconductivity have been surmounted one more time.

The materials and methods of preparation could not be more different — the one common feature is that they are still copper oxides. The first report⁴, from a Paris CNRS laboratory, describes the layer-by-layer deposition by molecular beam epitaxy of an eightfold sequence of CuO sheets interrupted by BiO intergrowths, a segment of what is commonly called the 'infinite layer' compound⁶. They observe a sharp drop in resistivity near 250 K in one sample, which seems to show a critical-current-induced shift in T_c when sample current is increased from 10^{-9} to 10^{-7} A. (The transport measurements are unorthodox; it seems that the current-voltage characteristics at various temperatures were used to obtain the resistivity plot.) These extraordinarily small currents are made necessary by the high sample resistance, which, although not explicitly stated, can reasonably be

estimated to be as much as 3 M Ω just before the resistivity drop. I have sometimes observed just such a drop when the sample resistance became large enough with decreasing temperature to saturate the internal impedance of the current source⁷. Unfortunately, the paper gives almost no detail on how the transport measurement was made. The magnetization data also leave open questions not addressed in the paper. The signals are extremely weak and the apparent diamagnetism commences at 300 K, about 40 K above the resistive onset, a most unusual result even for a filamentary superconductor. My overall concern is the paucity of routine experimental details and vagueness as to reproducibility.

The second report⁵ is from an experienced and well-known group based in Grenoble. Their samples were of the Hg-1223 compound (superconducting at 133 K)⁸ and its extensions to four and five copper oxide layers, bulk-synthesized under an ambient oxygen pressure of 18,000 atm, a preparative technique now becoming widely used. Only resistivity data are given in the paper; however, the experimental conditions are more thoroughly addressed than was the case for the Paris workers. No magnetization plots are given, just a few rather nebulous statements about small diamagnetic shifts, suggesting a superconducting volume fraction of about 0.01 per cent, which are not connected in an obvious way with the samples on which the transport data were taken.

The most provocative evidence for an extraordinary transition temperature is their observation of a sharp drop to zero resistance at 235 K. This is reported in a curious way, implying that the sample was left for two days without remeasurement (they didn't actually go home for the weekend, did they?). When the measurement was repeated it was with 20 times more sample current and the drop did not reappear. A question I would like to have seen addressed is whether the authors monitored the quadrature phase signal during the original measurement. On occasion, I have seen the in-phase component of the sample voltage drop to zero in a.c. measurements when one of the contacts became suddenly capacitive as the temperature was lowered. Neither paper indicates that either the transport or magnetic measurements were reproducible without hysteresis over many temperature cycles, a test that most superconductivity experimentalist referees would demand. My final comment about the Grenoble resistivity data is that the 235 K drop is strikingly sharp and narrow for

filamentary superconductivity, exhibiting neither a quasifluctuation rounding at the onset nor a tail as zero resistance is approached, the latter invariably associated with filamentary and non-bulk behaviour.

The years following the discovery of Bednorz and Müller gave rise to many sightings of Unidentified Superconducting Objects (USOs, an acronym which in Japanese means, I'm told, politely, 'error'). How then is one to judge reports of transient and ephemeral evidence suggestive of superconductivity? It is difficult. After all, in general there is no known theoretical reason why superconductivity cannot exist at almost any temperature: T_c in a neutron star is 10^9 K. For a given observation, the prime requirement has to be reproducibility, both in new sample preparation and remeasurement of original samples. For every association I have had with filamentary superconductors — from (SN), to the original ambient-pressure Bechgaard salts, the early formulations of the high- T_c materials, the observation of superconductivity in undoped La_2CuO_4 and the glimmer of 125 K in the thallium copper oxides — rapid confirmation was achieved by groups equipped to do so, once all conditions of preparation and experimental technique were known.

Are the French scientists thus expanding the revolution begun in 1986, or are they, more in the spirit of the 1825 uprising against Nicholas I, mere Decembrists? We should know the answer anon. Both groups claim they have reproduced their results in several samples. Perhaps even as we go to press, they will be able to obtain confirmation by distributing these samples to colleagues with the ability to repeat their measurements independently.

Finally, in expressing caution, and yes, perhaps scepticism, about these reports, we must be careful not to discourage the continuing quest for new superconductors, for, to paraphrase the now famous opening line of the Bednorz-Müller paper, this search is by its very nature empirical, with few rules to guide us. But along our journey it is prudent to remember Richard Feynman's admonition that in science it is easy to be fooled, and the easiest one to fool is yourself. □

Paul M. Grant is now in the Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94303, USA.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189 (1986).
2. Uchida, S., Takagi, H., Kitazawa, K. & Tanaka, S. *Jpn. J. appl. Phys.* **26**, L123 (1987).
3. Bednorz, J. G., Takashige, M. & Müller, K. A. *Europhys. Lett.* **3**, 379 (1987).
4. Laguerre, M. *et al. Science* **262**, 1850–1852 (1993).
5. Tholence, J.-L. *et al. Phys. Lett. A* **184**, 215–217 (1994).
6. Siegrist, T., Zahurak, S. M., Murphy, D. W. & Roth, R. S. *Nature* **334**, 231–232 (1988).
7. Ramirez, M. L., Vazquez, J. E. & Grant, P. M. *Internal IBM Rep. RJ6457* (28 September 1988).
8. Schilling, A., Cantoni, M., Guo, J. D. & Ott, H. R. *Nature* **362**, 56–58 (1993).

Deviants — or emissaries

Marvin Wickens and Kathy Takayama

RNA trespasses in what was once thought to be protein's province. The notion that RNAs can be enzymes, binding specifically to ligands, cofactors and substrates, is now commonplace; yet only a few years ago, these were the sacred acts of proteins. History may be about to repeat itself. Regulatory proteins bind to specific sequences in the genes or messenger RNAs they control, and so determine how much a gene is expressed, in what cells, and when. But why should these regulators have to be protein? Why not RNA? We already know, in bacteria, of RNAs that can control gene expression through remarkably sophisticated mechanisms. Now, two reports in *Cell*^{1,2} not only identify a tiny, repressing RNA in animal cells, but also show that it acts upon a region of mRNA often thought to be barren and insignificant. Although this could be a rare, deviant case, there is the tantalizing possibility that a new family of regulatory RNAs awaits discovery.

The new work centres on the regulation of the *lin-14* gene of the worm, *Caenorhabditis elegans*. This gene, which encodes a nuclear protein, controls the timing of developmental events; it is required for early developmental programmes, and must be repressed for late programmes to begin¹⁻³. What concerns us here is not how *lin-14* works, which remains mysterious, but how it is repressed at later stages of development.

Two players are critical: a repressor encoded by the *lin-4* gene and a site in *lin-14* mRNA itself¹⁻³. The *lin-4* repressor prevents the production of *lin-14* protein

in later stages. It does so by preventing the expression of *lin-14* mRNA after the mRNA has been synthesized and processed. Even when *lin-14* is fully repressed, its mRNA is present; the protein is not². In mutants that lack *lin-4*, the *lin-14* protein is present constitutively and so causes abnormal reiterations of early developmental events¹⁻³.

Not surprisingly, specific sequences in *lin-14* mRNA are needed for this repression — remove them, and the mRNA is not repressed^{2,3}; transfer them to a reporter gene, and that reporter comes under the control of *lin-4*, turning off at late stages in development². Remarkably, the key sequences lie between the termination codon and the poly (A) tail, in the so-called 3' untranslated region (3'UTR). The observation that sequences outside the protein-coding portion of *lin-14* mRNA control its expression suggests that the *lin-4* product inhibits, directly or indirectly, translation of *lin-14* mRNA, rather than controls *lin-14* protein turnover.

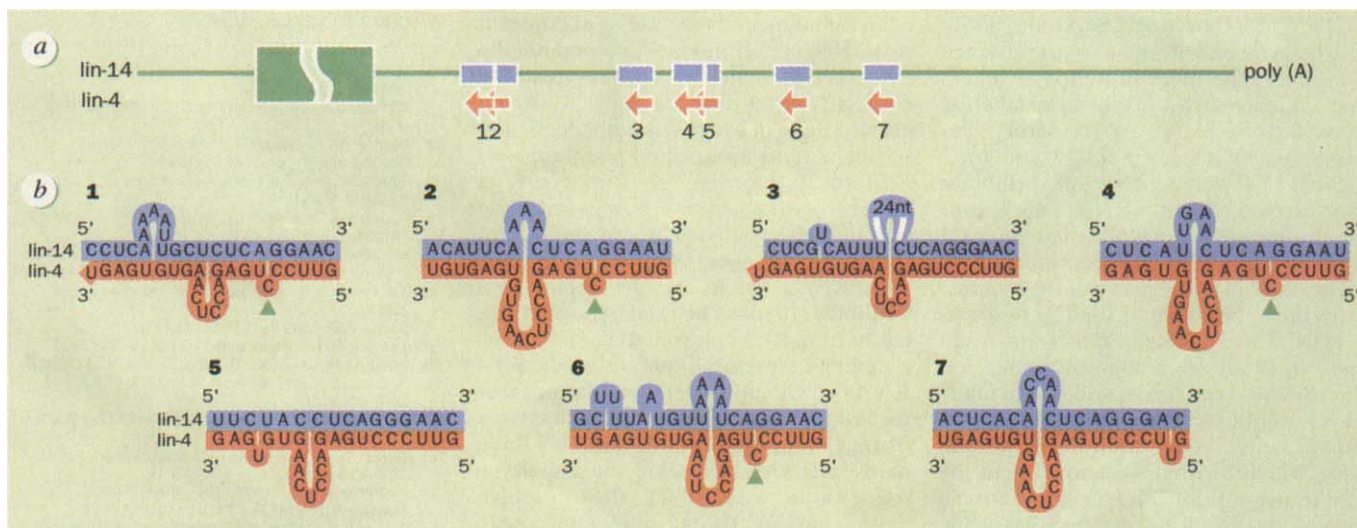
New regulatory elements in 3'UTRs are being discovered in profusion, not only in worms, but in flies, clams and vertebrates as well. This follows a long period in which 3'UTRs were viewed as barren and uninteresting. Understanding how elements in 3'UTRs work will require isolating the factors with which they interact, and the identification of *lin-4* as just such a molecule is therefore an exciting step. So exactly what protein does *lin-4* produce?

The answer seems to be, none. Instead, it produces only two RNAs, neither of

which appears to encode a protein¹. The most abundant species is positively tiny — only some 22 nucleotides long. The other, minor, RNA species extends about 40 nucleotides further at its 3' end. It seems very unlikely that the *lin-4* gene actually encodes a cryptic protein: although there are short open reading frames in the *lin-4* region, they are dispensable for repression and are not conserved between species. On the other hand, the nucleotide sequence in the region encoding the short RNAs is highly conserved, and even a single base change within it disrupts *lin-4* function¹. So it seems that the *lin-4* repressor is in fact RNA, not protein.

This can neatly explain how the key sequences in the *lin-14* 3'UTR are recognized (see figure)^{1,2}. The short *lin-4* RNA is complementary to a sequence reiterated imperfectly seven times in the 3'UTR of *lin-14*. On paper, seven distinct *lin-4*/*lin-14* RNA duplexes can be formed. In each, a base-paired stem forms between the 5' end of *lin-4* RNA and the conserved 3' end of the element, and a similar region of *lin-4* RNA (containing the sequence ACCUC) is looped out of the hybrid. A bulged C residue, present in four of the structures, may be crucial, because a mutation that substitutes a U eliminates *lin-4* function¹. It is unclear which, or how many, of the *lin-14* elements must be occupied to repress. The evolutionary conservation of the regions which pair, as well as the position of mutations in both *lin-4* and *lin-14*, show that these regions are critical, but do not themselves provide strong evidence for base pairing. For that, analysis of compensatory changes in *lin-4* and *lin-14* RNAs, or evidence of evolutionary covariation, will be required.

Although base-pairing is likely, it may not be the whole story. Genes carrying all



Proposed base-pairing between *lin-4* RNA and elements in the *lin-14* 3'UTR. *a*, *lin-14* mRNA, with the protein-coding region indicated by a large, broken box. Regions of the *lin-14* 3'UTR which potentially can pair with *lin-4* RNA are in blue, with their relative positions drawn to scale. *lin-4* RNA is in red. *b*, Hypothetical base-pairing schemes;

base-pairing has not been demonstrated experimentally (see text). The arrow indicates the C in *lin-4* which, when changed to U, prevents repression of *lin-14* mRNA. The termini of *lin-4* RNA have been deduced using DNA probes, rather than by direct analysis of the RNA, and so are imprecise. (Figure derived from refs 1, 2 and 8.)

seven of the *lin-14* repeats but lacking other conserved portions of the 3'UTR yield intermediate levels of repression, implying that sequences outside the repeats are important². These additional sequences could facilitate pairing between *lin-14* and *lin-4* RNAs by promoting conformational changes in the RNAs, as can occur in interactions between sense and antisense RNAs in prokaryotes⁴. Alternatively, the sequences could participate in a separate, *lin-4*-independent repressing mechanism; indeed, animals without *lin-4* still show a small amount of repressing activity².

Reiterated elements in 3'UTRs may be rheostats through which an mRNA's output can be modulated. In *lin-14*, mutant 3'UTRs carrying only three of the complementary repeats (3, 4 and 5 in the figure) still repress, though not as well as intact 3'UTRs (ref. 2). The same is true in *tra-2* mRNA of *C. elegans*, in which the 3'UTR negatively regulates expression through two tandemly repeated copies of a 28-nucleotide sequence. Mutants carrying only one element repress less well than a wild-type 3'UTR, but better than a 3'UTR with no elements at all⁵. These experiments suggest that partial occupancy of reiterated elements could provide a way to regulate expression incrementally.

Analysis of three genes in *C. elegans* — *tra-2* (ref. 5), *fem-3* (ref. 6) and *lin-14* (ref. 2) — suggests that regulatory elements in 3'UTRs may often evade genetic detection. The 3'UTRs of these three mRNAs contain negative regulatory elements that were first identified genetically, through strong selection for mutants that mis-expressed the gene. Such alleles are one to two orders of magnitude rarer than simple loss-of-function mutants, in which the gene's activity is destroyed. Thus comparable regulatory elements in the 3'UTRs of other mRNAs could easily have been missed. Perhaps, in the absence of mutants, the mere presence of reiterated sequences in a 3'UTR can be taken as prima facie evidence of regulatory elements: the 3'UTRs of *lin-14* and *tra-2* mRNAs, as well as that of vertebrate transferrin receptor mRNA⁷, each contain reiterated elements that control expression at a post-transcriptional level.

The work in *C. elegans* clearly demonstrates the importance of 3'UTRs in somatic cells. The 3'UTRs of *lin-14*, *tra-2* and *fem-3* mRNAs are recognized in several differentiated cell types, and unlike many 3'UTR elements previously identified in somatic cells⁸ do not control mRNA stability. Much of what we know about the importance of 3'UTRs in translational control and RNA localization has come from studies of mRNAs in eggs and early embryos^{9,10}. This may just be historical accident; or, more likely, it reflects the fact that post-transcriptional regulatory mechanisms are particularly common dur-

ing early development, before transcription of the zygotic genome begins.

How does *lin-4* RNA prevent expression of *lin-14* mRNA? In prokaryotes, antisense RNAs can repress translation, cause premature termination, or lead to cleavage by nucleases^{11–13}. In multicellular organisms, although the documented examples of naturally occurring antisense control are few, destruction of mRNAs after pairing to an antisense RNA may be common^{12,13}. In contrast, *lin-4* RNA does not appear to control *lin-14* mRNA's stability, but rather its translation. Although 3'UTRs have been implicated in translational control of a variety of mRNAs, the mechanisms are not yet clear enough to provide much guidance. Everything that a protein has been suggested to do by binding to a 3'UTR — whether it be regulating poly (A) length, controlling associations with the cytoskeleton, hiding mRNAs from the translation apparatus, or positioning them in the cell — an RNA could do too.

Could there be other short RNAs, as yet undiscovered, each regulating a different mRNA? More to the point, how might we find such RNAs if they do exist? Identifying short regulatory RNAs by classical genetics clearly is possible — it was done with *lin-4* after all — but it may require either extensive mutant hunts or good luck. In *lin-4*, loss-of-function mutants were isolated with a frequency roughly ten times less than for mutations that knock out typical protein-coding genes in *C. elegans*. Presumably the short length of the RNA reduces the chances of obtaining a mutation, as does the fact that many mutations may not sufficiently destabilize the proposed *lin-4/lin-14* RNA-RNA duplexes. To further complicate matters, the *lin-4* RNA sequence seems to lie within the intron of an mRNA, as has been found recently with several different small RNAs¹⁴. If this were commonly the case, then large genetic alterations would often affect not only the small RNA but also the gene in which it is embedded, and so might yield unexpected phenotypes.

From a biochemical standpoint, a search for short, rare RNAs without an assay for their activity or structure seems futile. The discovery of unusual 5' termini on snRNAs led to the development of antibodies that can be used to purify them easily from a complex mixture. Similarly, a detailed structural analysis of the *lin-4* RNAs might ultimately enable a search for related molecules. With the discovery of regulatory elements in the 3'UTRs of many mRNAs has come an assault on the factors with which they interact, employing gel shifts, ultraviolet cross-linking and screens of expression libraries. It might be wise to look specifically for *trans*-acting RNAs, as well as proteins.

Proteins may yet emerge that are required to control *lin-14* mRNA through its

3'UTR. It is unlikely that there is a protein whose sole function is identical to that of *lin-4*, as no other mutants have been identified with the same phenotype as *lin-4*. Nonetheless, a protein could still be directly involved — perhaps recognizing the duplexes, or facilitating pairing between the two RNAs, as occurs in other systems⁴ — but be required for other processes as well; if so, mutants might well have unexpected (and perhaps lethal) phenotypes. Surely additional factors are involved in the temporal control of *lin-4*, which leads to repression of *lin-14* only after a specific time during development. One exciting possibility is that it is not the abundance of *lin-4* RNA, but rather its ability to repress, that is turned on at a specific time. If so, then the *lin-4/lin-14* system could yield new modifiers of RNA-RNA duplexes, accelerators of RNA-RNA annealing, RNA helicases or other proteinaceous gadgets that act on RNA.

Odd, enigmatic RNAs lurk in the literature. There are RNAs that are polyadenylated and spliced like mRNAs, yet seem not to be translated^{15,16}. One, a 15-kilobase transcript of the *Xist* gene, has been implicated in chromosome condensation¹⁵, while another, H19, can act as a tumour suppressor¹⁷. Sea-urchin eggs contain a family of comparable RNAs of unknown function^{18,19}. Are these RNAs all grotesque deviants, one-of-a-kind aberrations, like characters in a Fellini film? Maybe. But pay close attention to them. They may turn out not to be oddities after all, but instead to have been our first emissaries from an unexplored and vast RNA world. □

Marvin Wickens and Kathy Takayama are in the Department of Biochemistry, 420 Henry Mall, University of Wisconsin, Madison, Wisconsin 53706, USA.

1. Lee, R. C., Feinbaum, R. L. & Ambros, V. *Cell* **75**, 843–854 (1993).
2. Wightman, B., Ha, J. & Ruvkun, G. *Cell* **75**, 855–862 (1993).
3. Ruvkun, G., Wightman, B., Burglin, T. & Arasu, P. *Development Suppl.* **1**, 47–54 (1991).
4. Eguchi, Y., Itoh, T. & Tomizawa, J. *A. Rev. Biochem.* **60**, 631–652 (1991).
5. Goodwin, E. B., Okkema, P. G., Evans, T. C. & Kimble, J. *Cell* **75**, 329–339 (1993).
6. Ahinger, J. & Kimble, J. *Nature* **340**, 346–348 (1991).
7. Harford, J. B. in *Control of mRNA Stability* (eds Belasco, J. & Braverman, G.) 239–266 (Academic, San Diego, 1993).
8. Sachs, A. B. *Cell* **74**, 413–421 (1993).
9. Jackson, R. J. & Standart, N. *Cell* **62**, 15–24 (1990).
10. Wickens, M. (ed.) *Sem. dev. Biol.* **3**, 363–433 (1992).
11. Simons, R. W. & Kleckner, N. A. *Rev. Genet.* **22**, 567–600 (1988).
12. Takayama, K. M. & Inouye, M. *Crit. Rev. Biochem.* **25**, 155–184 (1990).
13. Nellen, W. & Lichenstein, C. *Trends biochem. Sci.* **18**, 419–423 (1993).
14. Sollner-Webb, B. *Cell* **75**, 403–405 (1993).
15. Brockdorff, N. *et al.* *Cell* **71**, 515–526 (1992).
16. Brannan, C. I., Dees, E. C., Ingram, R. S. & Tilghman, S. M. *Molec. cell. Biol.* **10**, 28–36 (1990).
17. Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E. & Tycko, B. *Nature* **365**, 764–767 (1993).
18. Calzone, F. J., Lee, J. J., Le, N., Britten, R. J. & Davidson, E. H. *Genes Dev.* **2**, 305–318 (1988).
19. Nemer, M., Bai, G. & Stuebgen, E. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10851–10855 (1993).

New strings to the optical bow

R. F. Shi and A. F. Garito

NONLINEAR optics is the basis of all the fledgling photonics technologies, in which light works with, or even replaces, electrons in applications traditionally carried out by microelectronics. On page 47 of this issue, Kuwata-Gonokami, Peyghambarian and co-workers¹ add an entirely new form of nonlinear optical excitation to the list of those already observed — exciton strings, observed in an organic molecular solid of anthracene and pyromellitic dianhydride (anthracene-PMDA; see Fig. 1).

The prime candidates for technologically useful nonlinear materials — those whose refractive indices or absorption characteristics vary with light intensity — are high-performance conjugated organics and polymers², which show a remarkable range of ultra-fast, highly nonlinear optical excitations.

The origin of the outstanding nonlinear optical properties of organic systems is the relative ease with which light affects π -electron motions^{3,4}. The π -electron paths, or orbitals, extend over long distances, spanning an entire molecule, or even a

along — coulombically attracted to the electron — is the oppositely charged hole in S_0 . Bound together, they form a moving electron-hole pair, called an exciton. So much has been known for about 30 years, and researchers can directly detect single excitons in an organic, or polymeric, system by illuminating the system with intense light, then looking for characteristic absorption peaks.

The organic solid that Kuwata-Gonokami, Peyghambarian and co-workers used for observing excitons is a one-dimensional charge-transfer complex⁵. They chose the complex for their experiments because S_0 is located in a linear stack of molecules on a donor site (D, the anthracene) while S_1 lies on an immediately adjacent acceptor site (A, the PMDA) in the same stack. In the measurements, the individual excitons created by the incident light are confined to move only in a straight line along the stack.

The effect of this one-dimensional confinement is profound. In moving up and down a molecular stack, the excitons cannot easily avoid one another; they bump together, spontaneously forming pairs, triplets and longer assemblies called biexcitons, triexcitons and longer n -strings. A schematic illustration of a bound biexciton is presented in Fig. 2. The exciton n -strings become more numerous and easier to observe as more and more individual excitons are created with increased incident light intensity, because the otherwise faint absorption peaks now become pronounced. The entire process is nonlinear, as the creation and number of exciton n -strings depend in a nonlinear way on the light intensity. Further, in describing the pulling apart of an exciton along a stack, the authors show that the stability of one string (say, $n = 2$) depends on the existence of the next longer string ($n = 3$). This general condition reflects the nonlinear behaviour that is inherent to exciton n -strings.

The exciton n -strings join other excitations in condensed phases classified as collective excitations⁶. The list is long: examples include plasmons, which are back-and-forth oscillations of all the con-

duction electrons together in a metal, and rotons, which are associated with vortex motions in superfluid helium. In another example, electron-hole liquids can form in semiconductors when exciton concentrations are allowed to get so high that the excitons literally condense into separate electron-and-hole droplets which then behave like ordinary metals. The observation and properties of such collective excitations, in general, have been the

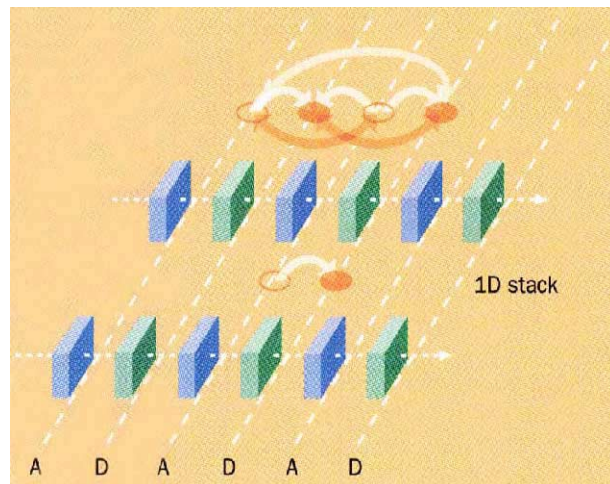


FIG. 2 Schematic illustration of the stability of exciton n -strings. The green and blue sites are donor (D) and acceptor (A) molecular sites, respectively. The white arrows indicate electron-hole Coulomb attractions (filled circles are electrons, open circles are holes). Orange arrows indicate electron-hole Coulomb repulsions. A single exciton is stabilized by the electron-hole attraction, and the biexciton is stable because the attractions outweigh repulsions. The stability of larger n -strings can be understood in a similar way.

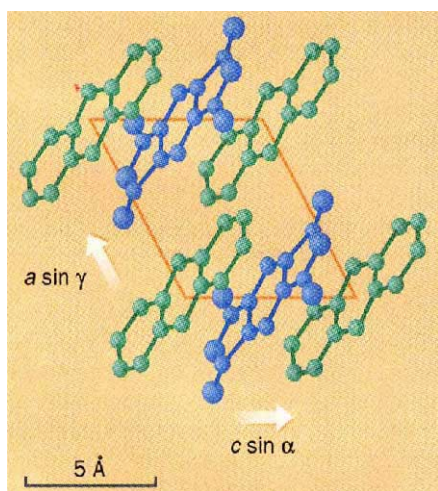


FIG. 1 Triclinic crystal structure of the quasi-one-dimensional charge-transfer complex anthracene-PMDA with anthracene donors (D) in green and PMDA acceptors (A) in blue. (Adapted from ref. 5.)

macroscopic solid. Delocalized in this way, the π -electrons are less tightly bound to the positive atomic nuclei of the molecule or solid than electrons in single atoms. As a result, to change paths, or equivalently, to promote a π -electron from its natural ground-state orbital S_0 to a nearby excited-state orbital S_1 , requires much less light energy. In addition, when the electron is removed from the ground S_0 , a positively charged hole is formed and left behind in S_0 . Now when the π -electron moves in the excited state S_1 , tagging

central focus of solid-state and condensed-matter physics for more than half a century.

Kuwata-Gonokami and co-workers' observation of the exciton n -strings extends our understanding of nonlinear optical processes in organic systems and points to the tasks as yet undone, which range from experimental measurements on new materials to the continued development of a comprehensive many-body theoretical description of the excitations. Their experimental discovery and its theoretical confirmation are fine examples of the beauty of fundamental nonlinear optical physics and the benefits of international scientific cooperation. □

R. F. Shi and A. F. Garito are in the Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

1. Kuwata-Gonokami, M. *et al.* *Nature* **367**, 47–49 (1994).
2. Hornak, L. A. (ed.) *Polymers for Lightwave and Integrated Optics: Technology and Applications* (Dekker, New York, 1992).
3. Heflin, J. R. & Garito, A. F. *Nature* **356**, 192–193 (1992).
4. Dalton, L. R. *Nature* **359**, 269–270 (1992).
5. Boeyenes, J. C. A. & Herbststein, F. H. *J. phys. Chem.* **69**, 2160–2176 (1965).
6. Walker, C. T. & Slack, G. A. *Am. J. Phys.* **38**, 1380–1389 (1970).

A tale of two functions

Joel A. Huberman

LAST year's discovery¹ of a multi-protein complex capable of ATP-dependent binding to yeast (*Saccharomyces cerevisiae*) replication origins generated much excitement, because it offered hope that the long search for an initiator protein that could recognize and activate eukaryotic replication origins might be over. The excitement was tempered, however, by the realization that, before a role for this origin recognition complex (ORC) in the initiation of DNA replication could be considered proven, the evidence for specific origin binding¹ would have to be reinforced by genetic evidence. Untempered excitement is now possible: four papers combine to provide convincing genetic evidence that ORC is indeed required for DNA replication²⁻⁵. Three of them²⁻⁴ also describe results suggesting that ORC is involved in transcriptional repression, raising the question of how ORC participates in such apparently dissimilar functions.

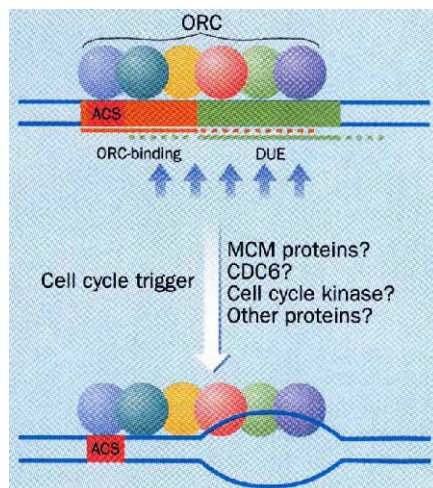
Hints about what ORC does during initiation of DNA replication are provided by what is known about interactions between ORC and yeast replication origins (see figure). The *cis*-acting sequences responsible for replication origins in yeast were first identified by the fact that plasmids containing them could replicate autonomously in yeast cells. They were therefore called autonomously replicating sequences, or ARS elements. Later experiments confirmed that ARS elements serve as replication origins in plasmids and showed that some also function as chromosomal replication origins⁶. Every ARS element has a close match to an 11 base-pair consensus sequence⁷ called the ARS consensus sequence, or ACS. In addition, a stretch of easily unwound DNA (a DNA unwinding element or DUE) has proved indispensable⁸.

ORC binds specifically to the ACS¹, and it protects a region (orange in the figure) from DNase I attack while at the same time stimulating attack at several specific sites (blue arrows in the figure), spaced at intervals of about 10 base pairs¹. The region containing DNase I-hypersensitive sites is also hypersensitive to copper phenanthroline, a probe for distorted DNA⁹. These observations suggest that the hypersensitive DNA may be wound around the outside of part of ORC and may be deformed. So a year ago it seemed that ORC might function, like the initiator proteins of prokaryotes, to bind to origins and distort them to permit (in cooperation with other proteins) the initiation of DNA replication^{1,9}.

With the intention of cloning the genes encoding the six subunits of ORC, Bell *et*

*al.*³ began to purify and determine partial amino-acid sequences of the individual subunits. They then found that the genes encoding two ORC subunits had already been cloned in other laboratories^{2,4,5}.

In one of them, Li and Herskowitz⁵ had searched for genes encoding proteins which, when fused to a transcriptional activator domain, would activate a gene whose promoter contained repeated



Interactions between ORC and a yeast replication origin. Spheres, the six ORC subunits; horizontal blue lines, the two strands of a DNA molecule; red box, the ACS; orange zone, the region (including the ACS) which is protected by ORC from DNase I; green zone, the DUE. As suggested by the dashed orange line, interactions between ORC and DNA (including induction of DNase I hypersensitive sites) extend rightward from the orange region. The dashed parts of the green line indicate that precise boundaries for the DUE have not yet been determined. The lower part of the figure symbolizes unwinding of the DNA, a key step in the initiation of replication.

copies of the ACS. They detected a gene encoding a polypeptide chain of M_r 50K. Comparison of the predicted amino-acid sequence of this polypeptide with partial amino-acid sequence information for ORC subunits³ showed that the cloned gene, *ORC6*, encodes the smallest of the six ORC subunits. Although the sequence of ORC6p (the protein product of *ORC6*) provides no information as to its function, *ORC6* is essential for cell viability, and its overexpression causes hypersensitivity to mutations in two other genes involved in initiation of DNA replication, *CDC6* (ref. 10) and *CDC46* (ref. 11). *CDC46*p belongs to the MCM family of proteins, members of which interact with each other¹¹, are important for DNA replication¹¹, and are conserved in eukaryotes ranging from yeasts to animals¹².

Elsewhere, Micklem *et al.*² and Foss *et*

*al.*⁴ had developed assays to search for ACS-binding proteins which depended on a completely different function of the ACS — its ability, with other *cis*-acting sequences and many *trans*-acting factors, to repress transcription of nearby genes¹³. The role of the ACS in transcriptional silencing had been established during investigations of the mechanism by which certain genes encoding yeast mating-type information are kept inactive. These investigations had demonstrated that repression of mating-type genes in yeast cells requires *cis*-acting elements called 'silencers'. Surprisingly, each of the four known silencers is associated with an ARS element¹⁴. Mutational analyses had revealed that the ACSs of the silencer-associated ARS elements are important for silencing¹³. Micklem *et al.*² and Foss *et al.*⁴ took advantage of that to assay for genes encoding proteins which contribute to silencing by binding to the ACS^{2,4}, and they independently discovered the same gene, with an open reading frame capable of encoding a 72K protein. It turned out that the gene, *ORC2*, encodes the second largest subunit of ORC, ORC2p.

Like ORC6p, ORC2p is essential for cell proliferation, but its amino-acid sequence reveals little about its function^{2,4}. However, the properties of cells bearing temperature-sensitive mutations in *ORC2* suggest that it is required for initiation of DNA replication. At restrictive temperatures, such cells can progress through most of the cell cycle but have difficulty entering S phase³. Even at permissive temperatures, such cells maintain ARS plasmids less efficiently than do wild-type cells^{2,4}.

Although biochemical and genetic observations consistently indicate that ORC has a central part in the initiation of DNA replication, the strength of ORC's DNase I footprint in permeabilized unsynchronized cells suggests that it is probably bound to ARS elements through most or all of the cell cycle⁹. Therefore binding of ORC to an origin is not likely to signal initiation of replication. Instead, some other cell-cycle-related event(s) must trigger initiation. ORC2p and ORC6p both contain potential phosphorylation sites^{2,4,5}, so phosphorylation by cell-cycle-regulated kinases may be a component of this trigger; cell-cycle-regulated alterations in the interactions between ORC and other proteins (perhaps CDC6p, CDC46p or other members of the MCM protein family) may also participate.

Details are still obscure, but it is now possible to imagine how ORC participates in the initiation of replication (see figure). How it functions in transcriptional silencing is much less clear.

One possibility is that local replication origin function is required to establish silencing. Here ORC would function in silencing as it does in replication: by

COULOMB FISSION

Caught in the act . . .



binding to the ACS, it would assist in activation of a replication origin. In favour of this possibility are the facts that DNA replication is required for establishment of silencing¹⁵, at least one silencer is a replication origin in the chromosome¹⁶ and all four silencers are ARS elements¹⁴ (replication origins in plasmids). An alternative is that ORC assists, with other proteins bound to other *cis*-acting silencer components, in tethering additional proteins which are necessary for silencing. In this case origin function would not be required. The requirement for DNA replication to establish silencing¹⁵ would be explained by the ability of replication to alter existing chromatin structure and permit formation of new chromatin structure. This possibility is supported by the observations that two silencers do not detectably function as replication origins in the chromosome¹⁷, and silencer mutants exist which are inactive as ARS elements but active as silencers¹⁸. Distinguishing between these possibilities will require improved definition of the *cis*-acting components of silencers and elucidation of the roles of the *trans*-acting factors required for silencer function.

A chapter in the story of ORC has ended, one in which it is plainly revealed that ORC has at least two functions. But another chapter now opens, in which we may hope to learn of the precise identification of the *cis*-acting sequences required for origin and silencer functions, and of the exact roles and cell-cycle regulation of the subunits of ORC, CDC6p, the MCM proteins, and other proteins required for replication and for silencing. □

Joel A. Huberman is in the Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14263, USA.

... but charges will be dropped. Above, an electrically charged heptane droplet ruptures to produce a fine spray of tiny offspring, each only about 10 μm in diameter. The large droplets are part of a stream expelled from a fine nozzle. They acquire charge as they leave the nozzle, which is held at a potential of a few kilovolts above a ground electrode, so coulombic repulsion between the charges acts against the surface tension holding the droplet together. If the surface charge is sufficiently high, the electrostatic charge overcomes surface tension and the drop bursts, as shown here.

The theory has been around for a while — the ever-curious Lord Rayleigh first looked into the problem in 1881, and derived an expression for when the droplet should disintegrate — but actually catching a drop mid-explosion has proved difficult. Alessandro Gomez and Keqi Tang snapped this one using a shadowgraph technique with 25-nanosecond flashes of light combined with a fast CCD (charge-coupled device) camera; as more than 10 thousand droplets per second are sprayed from the nozzle, there was a fair chance of photographing at least a few at the critical moment. In fact, they captured enough images to conclude that the common view of a violent, convulsive explosion is not borne out; instead, a well defined jet typically forms, as shown above. Their results are published this month in *Physics of Fluids*.

The Rayleigh limit on stability is reached when $q^2 = 8\pi^2\epsilon_0\gamma D^3$ where q is charge on the droplet, ϵ_0 the permittivity of the surrounding medium, γ the liquid surface tension and D the droplet

diameter. In real life, nothing is quite that simple, and effects such as aerodynamic distortion of the droplet as it falls, and the electric field between the nozzle and the ground electrode combine to reduce the limit (Gomez and Tang find disruption in general at charges of 70–80 per cent of the Rayleigh limit).

Rayleigh might have been surprised at the range of applications found for 'Coulomb fission'. Fine, evenly dispersed jets of liquid are needed for ink-jet printers, fuel combustion and, the application that sparked the experiments here, electrospray ionization as a basis for mass spectrometry of large, delicate biological molecules. John Fenn and collaborators at Yale University first demonstrated the principle in 1984: the species to be analysed are dissolved in a carrier liquid and electrostatically sprayed into fine droplets, divided by Coulomb fission into ever finer droplets, until the radius of curvature is so small and the electric field at the surface so high that ions are desorbed into the gas phase. More established ionization techniques, such as fast-atom bombardment, chemical ionization or inductively coupled plasmas, produce molecular ions with just one or two charges; the common run of mass spectrometers can detect these low-charged ions at relative molecular masses up to about $M_r = 10,000$. The electrospray method can dump 30 to 50 charges on each molecule, permitting the detection and identification (give or take a hundred atomic mass units or so) of molecular masses up to about $M_r = 100,000$, so proteins and other biological molecules can readily be analysed.

Lindsay Matthews

- Bell, S. P. & Stillman, B. *Nature* **357**, 128–134 (1992).
- Micklem, G., Rowley, A., Harwood, J., Nasmyth, K. & Diffley, J. F. X. *Nature* **366**, 87–89 (1993).
- Bell, S. P., Kobayashi, R. & Stillman, B. *Science* **262**, 1844–1849 (1993).
- Foss, M., McNally, F. J., Laurenson, P. & Rine, J. *Science* **262**, 1838–1844 (1993).
- Li, J. J. & Herskowitz, I. *Science* **262**, 1870–1874 (1993).
- Fangman, W. L. & Brewer, B. J. *A. Rev. Cell Biol.* **7**, 375–402 (1991).
- Newlon, C. S. & Theis, J. F. *Curr. Opin. Genet. Dev.* **3**, 752–758 (1993).
- Huang, R.-Y. & Kowalski, D. *EMBO J.* **12**, 4521–4531 (1993).
- Diffley, J. F. X. & Cocker, J. H. *Nature* **357**, 169–172 (1992).
- Hogan, E. & Koshland, D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3098–3102 (1992).
- Chen, Y., Hennessy, K. M., Botstein, D. & Tye, B.-K. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10459–10463 (1992).
- Thömmes, P. et al. *Nucl. Acids Res.* **20**, 1069 (1992).
- Herskowitz, I., Rine, J. & Strathern, J. in *The Molecular Biology and Cellular Biology of the Yeast Saccharomyces: Gene Expression* Vol. 2 (eds Jones, E. W., Pringle, J. R. & Broach, J. R.) 583–656 (Cold Spring Harbor Laboratory Press, New York, 1992).
- Broach, J. R. et al. *Cold Spring Harb. Symp. quant. Biol.* **47**, 1165–1173 (1982).
- Miller, A. & Nasmyth, K. A. *Nature* **312**, 247–251 (1984).
- Rivier, D. H. & Rine, J. *Science* **256**, 659–663 (1992).
- Dubey, D. D. et al. *Molec. cell. Biol.* **11**, 5346 (1991).
- Mahoney, D. J. et al. *Genes Dev.* **5**, 605–615 (1991).

Marriage of the flytrap and the serpent

Bruce R. Conklin and Henry R. Bourne

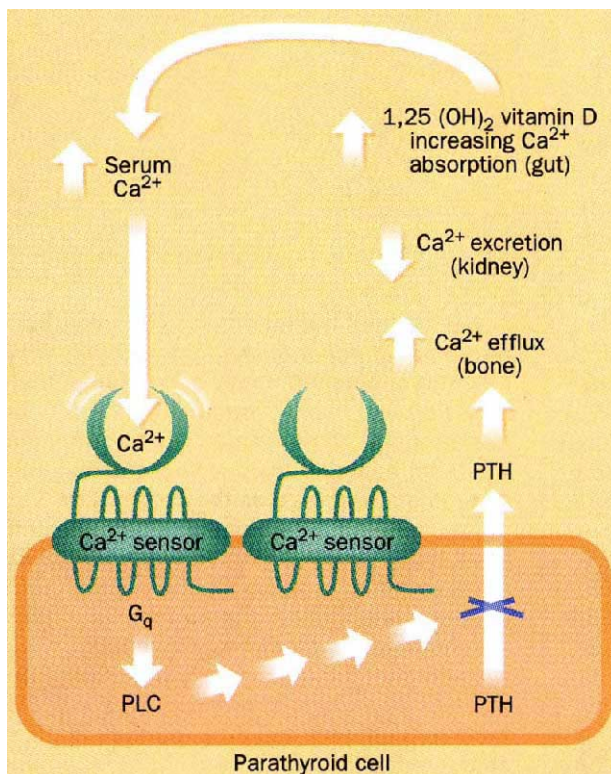
THE cloning of a sensor for extracellular calcium, described by Edward Brown and colleagues in *Nature* last month¹, reveals the evolutionary marriage of two protein structures that hail from different kingdoms. One works (and looks) like a Venus flytrap and belongs to a family of bacterial detectors for ions and nutrients. The other, a serpentine receptor coupled to G proteins, is found in myriad incarnations in eukaryotes, where it serves as a detector for light, pheromones, odours, hormones and neurotransmitters.

The newly discovered sensor regulates secretion of parathyroid hormone (PTH), the principal hormonal regulator of serum Ca^{2+} in mammals. Acting in bone, kidney and the gut, PTH raises serum Ca^{2+} (see figure). Appropriate regulation of PTH secretion requires that PTH-secreting cells in the parathyroid glands 'sense' ambient serum calcium². The newly cloned sensor responds to increased levels of extracellular Ca^{2+} by triggering a phospholipase-C-dependent pathway that tells the cell to decrease PTH secretion.

Brown and collaborators in the Seidman laboratory have moved fast to highlight the physiological significance of the sensor. In the latest issue of *Cell*³ they report the obvious gene-knockout observation — and in an increasingly popular animal model, people. Inheritance of one inactive calcium sensor gene causes familial hypocalcaemic hypercalcaemia (FHH), a mild heterozygous, dominant disorder characterized by modestly elevated serum Ca^{2+} , normal PTH levels and inappropriately low urine Ca^{2+} . Together, these findings suggest that patients with FHH have a raised 'set point' for extracellular Ca^{2+} (normally, increases in serum Ca^{2+} lead to decreased PTH levels and increased urinary Ca^{2+} excretion). Serum Ca^{2+} is much higher in the unfortunate infant who inherits two defective Ca^{2+} sensor genes; this homozygous, recessive disorder is lethal unless the parathyroid glands are surgically removed.

Identification of a calcium sensor sparks new hopes for finding the sensors of other extracellular ions. Which ion will next achieve the status of a hormone? Some of the candidates (Mg^{2+} , PO_4^{2-} , Na^+ , Cl^-

and H^+) are already known to activate second messenger systems associated with G protein-coupled receptors². Routine clinical measurements of ion concentrations have led to discovery of genetic disorders characterized by abnormal



The calcium sensor's role in parathyroid hormone (PTH) secretion and calcium homeostasis. The extracellular domain is shown closing upon binding to calcium, analogous to the action of a Venus flytrap. PLC, phospholipase C; the G protein is shown as G_q (the subclass of G protein thought to activate PLC).

extracellular concentrations of ions⁴ such as K^+ , PO_4^{2-} and Mg^{2+} . Inherited mutations of an ion receptor could account for either an increase or a decrease in the extracellular concentration of that ion, depending upon whether the mutation inactivates the receptor or makes it more than normally sensitive to its ligand (the second possibility is less likely, but has precedents among G protein-coupled receptors⁵).

The Ca^{2+} sensor is unusual among serpentine receptors because it has an enormous extracellular domain of 605 amino acids at the amino terminus (see figure). This domain probably serves as the actual Ca^{2+} detector, while the serpentine domain (composed of the seven membrane-spanning segments characteristic of G protein-coupled receptors) transmits the message across the membrane. Of the three point mutations in

FHH patients³, two are in the extracellular domain and one in a portion of the serpentine domain predicted to contact G proteins⁶. Similarities in amino-acid sequence suggest that the extracellular domain is a modified version of a large family of periplasmic binding proteins (PBPs) that are involved in bacterial nutrient uptake; these proteins bind ions and other nutrients in the periplasm, and deliver them to transporter proteins that move the nutrients across the plasma membrane⁷. Crystal structures of PBPs resemble a Venus flytrap, with two lobes that move like a pincer to envelop the ligand⁸.

The first described resemblance between PBPs and serpentine receptors involved the metabotropic glutamate receptor, whose extracellular ligand-binding domain has been modelled after the PBP structure⁹. Strikingly, in FHH patients both point mutations in the extracellular domain of the Ca^{2+} sensor align with the predicted ligand contact sites of the metabotropic glutamate receptor (refs 3 and 9; P. J. O'Hara, personal communication). Such mutations support the speculative idea that Ca^{2+} binding triggers the two lobes of the Ca^{2+} sensor's extracellular domain to bend towards one another, producing a large conformational change that in turn tells the serpentine domain to activate G proteins.

The calcium sensor's role in physiology is at long last emerging, and the sensor's identification illustrates the power of molecular biology to produce structural and functional insights by uniting disparate fields such as bacterial nutrition and clinical endocrinology. The first union of the flytrap and the serpent was probably something of a shotgun wedding. But one of their offspring has evolved into a precisely machined and vital component of normal homeostasis in mammals. □

Bruce R. Conklin and Henry R. Bourne are in the Departments of Pharmacology and Medicine, and the Cardiovascular Research Institute, University of California, San Francisco, California 94143, USA.

1. Brown, E. M. *et al.* *Nature* **366**, 575–580 (1993).
2. Brown, E. M. *Physiol. Rev.* **71**, 371–411 (1991).
3. Pollak, M. R. *et al.* *Cell* **75**, 1297–1304 (1993).
4. McKusick, V. A. & Francomano, C. A. *Mendelian Inheritance in Man* (Johns Hopkins University Press, Baltimore, 1992).
5. Lefkowitz, R. J. *Nature* **365**, 603–604 (1993).
6. Kobilka, B. A. *Rev. Neurosci.* **15**, 87–114 (1992).
7. Adams, M. D. & Oxender, D. L. *J. biol. Chem.* **264**, 15739–15742 (1989).
8. Oh, B.-H. *et al.* *J. biol. Chem.* **268**, 11348–11355 (1993).
9. O'Hara, P. J. *et al.* *Neuron* **11**, 41–52 (1993).

Ice-age tropics revisited

David M. Anderson and Robert S. Webb

MODELS of future climate should be able to simulate the past when boundary conditions were different, a critical test of model reliability. But there has long been a problem with simulating temperatures worldwide during the last glacial period. Models that use the sea surface temperatures derived by CLIMAP (the Climate Long-range Investigation and Mapping programme) as boundary conditions stubbornly refuse to fit palaeoclimate evidence of lowered snowlines and vegetation change — the modelled tropics are too warm. At the American Geophysical

Union Meeting in San Francisco, on 10–11 December 1993, the message was that the sea surface temperatures in the tropics were indeed 5 °C lower than today. Can CLIMAP be wrong?

The evidence comes from corals around Barbados (T. P. Guilderson, Lamont-Doherty Earth Observatory), and matches the cooling (4–7 °C) estimated from terrestrial evidence of lower snowlines and changes in ancient patterns of vegetation at mid- to high elevations. Other marine proxies have indicated less than 2 °C cooling of the tropical ocean, creat-

ing a discrepancy between marine and terrestrial indicators of tropical climate too large to be attributed to estimation errors.

One way to explain the discrepancy would be if the lapse rates (the changes of air temperature with height, assumed to be the same as today) had in fact been steeper in the past. At the conference, this idea was effectively scotched, at least for southwestern North America. The noble gas content of glacial-age aquifers from heights of 200 and 2,000 m indicates that lapse rates in this region were unchanged (M. Stute, Lamont-Doherty Earth Observatory).

The magnitude of change in the tropics is more than a curiosity, for it provides an essential measure of the sensitivity of the

OBITUARY

Lewis Thomas (1913–93)

LEWIS Thomas, the physician, scientist and educator who died on 3 December last year, was one of the foremost essayists of our time. His column, *Notes of a Biology Watcher*, serialized in *The New England Journal of Medicine*, and published as *The Lives of a Cell* in 1974, brought him acclaim far beyond the medical profession and earned a National Book Award in Arts and Letters. Thomas's long career brilliantly combined the two cultures of art and science.

He was born in Flushing, then a quiet suburb of New York City, where he often accompanied his father, a physician, on his daily rounds. It was here that he learned the art of medicine — but of science, as he so often noted, there was precious little. A graduate of Princeton University (where, in 1986, a biomedical laboratory was named after him), he went on to Harvard Medical School and left there in 1937 eager for a career in biomedical science. When war was declared against Japan, Thomas joined The Rockefeller Hospital Naval Unit, served in Guam and later landed on Okinawa clutching a box of fifty mice that would help identify the Japanese B encephalitis virus present on the island. Returning home, he embarked on a medical and scientific career that was as dazzling as it was diverse.

At various times, Lewis Thomas held chairs of paediatrics, pathology, microbiology and medicine. To him they were not different disciplines but all wonderful opportunities to explore the pathogenesis of disease and the mysteries of human biology. Wherever he went and however burdened by administrative responsibilities, whether as dean of the medical school at Yale, or as president of the Memorial Sloan-Kettering Cancer Center, he always

had his own small laboratory where he could dream and puzzle over the significance of bacterial endotoxins and of mycoplasma, problems in microbiology and immunology that particularly interested him. Perhaps his most imaginative conceptual



Lewis Thomas, a poet laureate and statesman of science.

leap was that of immune surveillance, an idea that was popularized by Sir MacFarlane Burnet. It is not fully elaborated, but is central to our understanding of many immunological diseases including cancer and AIDS.

In 1954, Thomas became professor of pathology at New York University School of Medicine, and built an outstanding department there. Emphasizing both research and training, he inspired many younger associates who would rise to positions of great distinction (one of them, Baruj Benacerraf, would receive a Nobel prize). Thomas served on many academic and government

advisory committees, including the President's Scientific Advisory Committee, and to all this work he brought a scholarly viewpoint, a perceptive eye and immense goodwill.

As a spokesman for science, Thomas also had the rare ability to transmit the spirit and wonder of science to those for whom it is not a primary concern. A supporter of the Gaia hypothesis, it was his conviction that the Earth is ultimately sustained by coherence, cooperation and complex self-regulatory systems — as he put it, "the earth is a sort of immense organism".

As well as *The Lives of a Cell*, translated into eleven languages, his published essays include *Medusa and the Snail* (1979), *The Youngest Science* (1983) and *Late Night Thoughts on Listening to Mahler's Ninth Symphony* (1983). His last book, published in 1992, was entitled *The Fragile Species*. His books and graceful writings have enriched the lives of countless people, young and old, for they are enchanting essays that not only inform us but make us more optimistic about the world. Thomas was a member of the National Academy of Sciences, The American Philosophical Society and the American Academy of Arts and Sciences, and received more than 20 honorary degrees, not only in science but in law, literature and music.

In 1993, The Rockefeller University established The Lewis Thomas Award to "honor the scientist as poet". Small wonder that Lewis Thomas, a poet laureate and statesman of science, should receive the first award.

Alexander G. Bearn

Alexander G. Bearn is at The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

tropics to past (and future) climate change, and an important test for models constructed to predict future climate. General circulation models that prescribe sea surface temperatures are unable to simulate the much colder temperatures observed near the poles and at high elevations unless low-latitude sea temperatures are colder¹.

Previous estimates of glacial sea surface temperatures are not easily dismissed because they are derived from multiple proxy sources from each of the tropical oceans (several are summarized in the table). CLIMAP's conclusion² in 1981 that there had been little temperature change (2 °C) was based on fossils of several plankton groups (radiolarians, coccolithophores, diatoms and foraminif-

Method	Change in SST	Uncertainty
Coral Sr/Ca	-5	±0.5
Coral $\delta^{18}\text{O}$	-5	±0.5
CLIMAP	-2	±1.5
Foraminifer $\delta^{18}\text{O}$	-2	+1
Algae alkenones	-1	±0.5

Differences between glacial and present sea surface temperature (SST, in °C) and a rough estimate of the between-sample variability for several methods.

ers) whose distributions in the tropical ocean have not changed since glacial times. Judging from their modern distribution, a cooling of 5 °C would bring an entirely different assemblage of species into the tropical ocean surrounding Barbados. Geochemical evidence associated with the plankton also indicate little change. The temperature-dependent fractionation of oxygen isotopes, $^{18}\text{O}/^{16}\text{O}$, preserved in foraminifer calcite indicates less than 2 °C cooling, and new results based on the temperature-dependent saturation of long-chain alkenones from algae indicate less than 1 °C cooling in the western tropical Pacific (N. Ohkouchi, University of Tokyo; K. Kawamura, Tokyo Metropolitan University).

Guilderson and colleagues base their results on the temperature-dependent way in which strontium and calcium, and the different oxygen isotopes, are incorporated into the aragonitic coral skeleton. With distribution coefficients determined from modern corals growing over several seasons in water of known temperature, both approaches yield a temperature decrease of 5 °C for corals with a radiocarbon age of 15–16 kyr (see table). Good agreement has been found between the seasonal cycle in sea surface temperature and Sr/Ca-based estimates for Pacific corals, providing an impressive test of the Sr/Ca approach³. Assuming that ocean Sr/Ca ratios were unchanged at 15–16 kyr ago (logical because of the long

residence time), the Sr/Ca technique depends only on the biologically mediated precipitation of aragonite.

The $^{18}\text{O}/^{16}\text{O}$ ratio is more complicated, being a function of temperature and the ambient composition of sea water. The change in $^{18}\text{O}/^{16}\text{O}$ in the Barbados coral on going from glacial to interglacial climate (2.3 parts per thousand) is particularly puzzling because it is 1‰ greater than the difference observed in foraminifer calcite obtained from sediment cores near Barbados (1.2–1.6‰). If it is attributed entirely to temperature, the difference is equivalent to 4 °C.

What is the source of the differences between temperatures derived from corals, foraminifers and prymniophyte algae? The answer may lie in ecological and physiological processes that are at present only partly understood. Foraminifer zooplankton can, under certain circumstances, grow at different depths and during different seasons, picking up a different or mixed environmental signal. Calcite added at the end of the organism's life or removed at the sea floor complicates the picture. In contrast the algae, producers of long-chain alkenones, have the ability to bloom rapidly under certain conditions and thus record a snapshot of the ocean environment over a narrow interval (J. Kennedy, Indiana University).

Corals might seem the ideal monitor, fixed in space and growing throughout the year in a way that can be determined from the annual growth bands. But calcification is mediated by kinetic effects, and modern corals are out of equilibrium with respect to seawater Sr/Ca and $^{18}\text{O}/^{16}\text{O}$, an effect that varies with growth rate and between species. At low growth rates, coral Sr/Ca and $^{18}\text{O}/^{16}\text{O}$ increase, giving the appearance of lower temperatures from both parameters^{4,5}. Guilderson and co-workers stressed that they were careful to avoid sections of coral that had different growth rates in order to minimize the effect of kinetic fractionation. Unless identified or removed, however, changes in the extent of disequilibrium complicate the signal.

To be certain that any one of these organisms is reliable will no doubt require an improved knowledge of the ecology and physiology of each of these natural recorders of ocean history. □

David M. Anderson and Robert S. Webb are at the NOAA Paleoclimatology Program, 325 Broadway, Boulder, Colorado 80303, USA.

1. Rind, D. & Peteet, D. *Quat. Res.* **24**, 1–22 (1985).
2. CLIMAP Project Members *Geol. Soc. Am. Map and Chart Series*, MC-36 (1981).
3. Beck, J. W. *et al. Science* **257**, 644–647 (1992).
4. McConnaughey, T. *Geochim. cosmochim. Acta* **53**, 151–162 (1989).
5. Devilliers, S., Shen, T. & Nelson, B. K. *Geochim. cosmochim. Acta* (in the press).

A personal view

THE doctrine of 'privacy of sensation' states that nobody can ever know what anyone else hears, or sees, or feels. Fortunately, wide areas of agreement can be reached in practice. Daedalus now plans a small step towards knowing what someone else sees.

The ophthalmologist scrutinizes your retina with a retinoscope, whose concave lens cancels the convex lens of your eye. DREADCO opticians are devising a small retinoscope mounted on a spectacle frame, from which a tiny fibre-optic stalk peers into the pupil of the eye from one side. The wearer can still see normally; the little stalk is out of focus and quite unobtrusive. The device sees not merely the retina itself, but the image of the scene projected on it. A fibre-optic cable conveys the image to a small belt-mounted video recorder.

You might think that eye movements would defeat this arrangement. But Daedalus argues that such movements merely shift the retina about beneath the static optical image. The constant movements will even help the system, by tending to average out the distracting background of retinal structure.

Once the image is on tape, the next step is to play it back up the fibre-optic cable, and project it back onto the subject's retina. With all the power of modern computer-video electronics to help him, he can then tweak the replay until it exactly resembles his original vision. Under his guidance, the computer will subtract the known retinal structure from the image, correct its optical distortions, and adjust its colours and intensities. Finally, he will announce that this, indeed, is what he originally saw. A replay onto a video screen will then reveal his vision to a wider audience, while the tweaking it needed will reveal his private optical distortions. Two sight recorders, one for each eye, could make his stereoscopic depth perception public as well.

Ophthalmologists, vision psychologists and students of optical illusions will be fascinated. But Daedalus sees wider commercial applications. DREADCO's personal sight recorder will completely take over the amateur photography and video market. Who will want to lug a heavy camera around when he can record anything he sees through his own eyes, and in stereoscopic depth as well? Crime and social deviancy will plummet when every stranger's eye may be recording the scene, ready for replay in the witness box. And a whole new field of nostalgia will open up. The fading photograph and family album will be replaced by the ability to see again the scenes of the distant past. David Jones

Antimatter and the Moon

SIR — The recent measurement of the shadow of the Moon in charged cosmic rays with energies in the 10 TeV energy region¹ shows a displacement of 0.15° from the Moon's position; as pointed out previously² the Earth–Moon system acts as an ion spectrometer and can be used to differentiate matter and antimatter in the cosmic radiation. The direction and magnitude of the shift in this Tibetan air-shower experiment is in agreement with the hypothesis that most of the cosmic radiation at this energy is composed of protons (matter). Although the cosmic radiation does show an anomalously high flux of anti-protons at energies of 1–10 GeV (ref. 3), there is no direct experimental evidence of the nature of the cosmic radiation at energies above 20 GeV; in principle all of this high energy component could be antimatter.

On the assumption that the Universe is baryon symmetric and that higher-energy cosmic rays are of extragalactic origin, Stecker and Wolfendale⁴ have shown that the observed ratio of $\bar{p}/p$ below 20 GeV can be explained; at energies of 10 GeV this ratio is 0.1%, and the prediction is that it rises with energy to as much as 20% above 10 TeV.

The significance of the Moon shadow detection is at the 7.1σ level (Fig. 1); the displacement of the centroid of the shadow image from the known position of the Moon is $(0.16 \pm 0.11 / -0.09)^\circ$ to the west and $(0.02 \pm 0.07 / -0.09)^\circ$ to the south. The calculated shift of the image by the geomagnetic field is estimated to be 0.15° to the west¹ for the mode energy of protons at the zenith angle of the observations.

If the cosmic radiation was composed

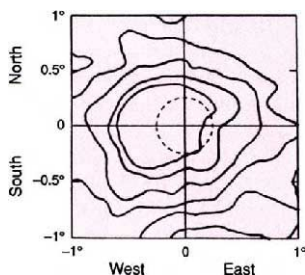


Fig. 1 Distribution of arrival directions of air showers in a box $2^\circ \times 2^\circ$ centred on the position of the Moon as recorded in the Tibet air shower experiment¹.

entirely of antiprotons, the shadow would be displaced to the east by the same amount; a mixed composition would result in an elongated or displaced image. We have simulated the Moon image for various mixtures of matter and antimatter

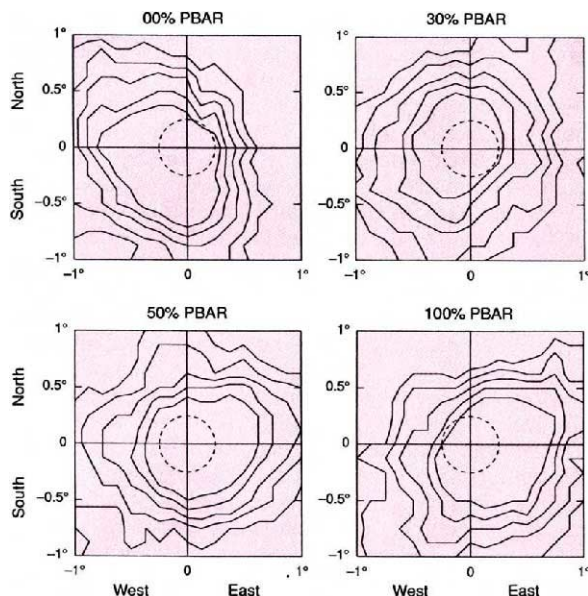


Fig. 2. Simulations of the arrival directions of a similar number of events assuming the composition $(\bar{p}/p + p)$ is 0, 30, 50 and 100%.

assuming a uniform magnetic field deflection of 30 mrad E^{-1} (TeV) and using a similar analysis procedure to that used in the Tibetan experiment (Fig. 2). The clear displacement of the measured image to the west (matter) and the lack of any significant elongation in the east–west direction agrees with the hypothesis that the bulk of the cosmic radiation at these energies is composed of protons. Although it is difficult to arrive at a quantitative limit without a detailed modelling of the Tibet experiment, a conservative figure (based on a visual comparison of Figs 1 and 2) is that the antimatter component is less than 30%; this limit is not sufficient to challenge the baryon symmetric model⁴ but does show that all the high-energy cosmic rays cannot be antimatter.

At energy thresholds of 1 TeV the offset is as much as 1.5° . More precise measurements of the moon shadow and hence of the $\bar{p}/p$ ratio in the 1–10 TeV region will soon be forthcoming from the

ARTEMIS² and CLUE⁵ experiments using the atmospheric Cherenkov technique in the ultraviolet.

M. Chantell

T. C. Weekes

Harvard-Smithsonian Center for Astrophysics,
Whipple Observatory,
PO Box 97, Amado,
Arizona 85645-0097, USA

X. Sarazin

M. Urban

LPNHE, Ecole Polytechnique/
CNRS-IN2P3,
91128 Palaiseau, France

Ancient rituals in Gabon

SIR — During a palaeontological survey of caves in Gabon, West Africa, in 1992, we discovered evidence of past human presence in Paouan cave, near Lastoursville¹, well beyond the limit of light penetration. A sondage in the sediments comprising the floor of a vast chamber revealed the presence of archaeological levels which contained abundant stone tools and charcoal. We examined samples of charcoal by microscope to determine the plant species present, and analysed others by the ^{14}C method to obtain evidence of the ages of the layers. The earlier of the two levels is $5,570 \pm 70$ years before present, the later is $4,000 \pm 70$ years before present.

The stone tools from the lower level consist of hammer stones, flakes, cores, blades and rarer triangular points worked in black to grey jasper and milky quartz. The assemblage as a whole resembles the Tschitolean industry which is widespread in Zaïre, Congo and Gabon². Both types of raw material consist of water-worn pebbles introduced into the dolomite cave by humans.

Among more than 80 charcoal fragments identified, only four species of plant are represented, which indicates that the choice of plants carried into the cave was highly selective, considering that the forest surrounding the cave is extremely diverse (more than 1,000 woody plant species grow in Gabon). The four plant species found in archaeological context in Paouan cave are well known to local inhabitants, all of them having magical-ritual properties. The tree *Pterocarpus tinctorius* yields a red powder used to colour the body, especially during ceremonies such as initiation rites. *Copaifera* resin (copal) is used to manufacture torches, the tree itself considered as the 'king of trees' by many people of the African tropical forest. *Combretum* yields extremely hard wood often used to make batons and clubs, the wood being said to impart "aggressive" qualities to the holder. Finally, the liana *Strophanthus* is used

1. Amenomori, M. et al. *Proc. 23rd Int. Cosmic Ray Conf.* **4**, 351–354 (Calgary, 1993).
2. Urban, M. et al. *Nucl. Phys. (Proc. Suppl. 14B)*, 223–236 (1990).
3. Stephens, S. A. & Golden, R. L. *Space Sci. Rev.* **46**, 31 (1987).
4. Stecker, F. W. & Wolfendale, A. W. *Nature* **309**, 37–38 (1984).
5. Appollinari, G. et al. *Nucl. Inst. Meth. A* **263**, 255 (1988).

by many African tribes as a source of arrow poison and as a tonicardiac during periods of violent exercise such as iron smelting or prolonged ceremonies.

The highly restricted diversity of plant species carried into Paouan cave, allied to their known properties and modern uses by tropical African tribes, indicate that we may have discovered a place in which magical-ritual practices were carried out around 5,600 years ago. The presence of worked stone and *Strophanthus* charcoal suggests that poison arrows were manufactured in the cave. Poison-arrow manufacture by many African tribes is carried out with elaborate ritual in secret places, such as caves, rock shelters or forest thickets, from which women are excluded. The choice of *Copaifera* resin as the preferred torch-making material, and of *P. tinctorius* accords with the view that Paouan cave was used for magical-ritual purposes.

It seems that by 5,600 years ago, some of the tribes of tropical Africa had mastered the art and phytochemistry of poison-arrow manufacture, and that they had developed elaborate rituals around this activity.

Richard Osisly

Paléogab, 4 rue Niger, Paris, France

Martin Pickford

Geological Survey of Namibia,
PO Box 2168, Windhoek, Namibia

Roger Dechamps

Royal African Museum,
BP 1980, Tervuren, Belgium

Michel Fontugne

Laboratoire Mixte CNRS-CEA,
Centre des Faibles Radio Activités,
91198 Gif-sur-Yvette, France

1. Delorme, G. *Spelunca* 4, 151–160 (1979).

2. Van Noten, F. (ed.) in *The Archaeology of Central Africa* 27–40 (Academische, Gruz, 1982).

Parsimony or statistics?

SIR — In the light of Stewart's Review article¹ on parsimony, I would like to comment on the use of this technique with molecular sequence data. Phylogenetic inference, no matter which method is applied, hinges on the tenet that the method used takes into account the mode of evolution of the molecule analysed. To infer phylogenies from DNA sequence data, there must be a mathematical description of the evolutionary process — a stochastic model.

Such models have been incorporated into statistical methods of phylogenetic inference, such as maximum likelihood^{2,3} or corrections of pairwise distances⁴. Methods with general models have wide applicability, allowing analysis of sequences that have undergone few or many substitutions, that evolve by a transition

bias, whose positions evolve at different rates, or whose base composition is biased. By contrast, parsimony's implicit stochastic model is highly restrictive. For example, it requires that each nucleotide position has a negligible probability of having changed more than once⁵. It generally is not known *a priori* if evolution generated one's data in accord with this rule. Parsimony has furthermore been shown to be statistically inconsistent (to arrive at the wrong answer at statistical significance) in a wide range of relatively realistic schemes^{6–8}. Parsimony may have its limited set of applications, but for the average phylogeneticist it is not the method of choice.

Further, Stewart states that "[Distance] methods ignore the possibility that apparent overall similarity and true evolutionary relationship are not necessarily the same thing." This statement is incorrect. When properly applied, certain distance methods are powerful methods of phylogenetic inference that make less restrictive assumptions about the evolutionary process than does parsimony. The neighbour-joining method⁹, for example, with properly corrected pairwise distances, is a reliable method of phylogenetic inference even in the absence of a molecular clock⁸.

For nonspecialists, the following may serve as a brief statistical guide for conducting or evaluating phylogenetic sequence analyses. (1) Are the assumptions of the method used met by the mode of evolution of the molecule analysed? Does the method's explicit or implicit stochastic model take account of biased base composition, of different rates of transitions and transversions, or of variations among sites in the rate of evolution, for example? If an analysis was done with parsimony, is its use explicitly justified? (2) Were enough data collected to allow evaluation of multiple internal nodes (those nodes that determine the exact branching pattern of the tree)? (3) Are the conclusions based on significance obtained with a statistical method of phylogenetic inference, such as maximum likelihood, neighbour-joining with properly corrected distances, or invariant analysis? Application of statistical methods of phylogenetic inference avoids many of the pitfalls of parsimony, but has all of its powers.

Arend Sidow

Department of Molecular Cell Biology,
University of California,
Berkeley, California 94720, USA

STEWART REPLIES — The purpose of my invited Review¹ was to discuss the powers and pitfalls of parsimony analysis in modern molecular evolutionary biology, not to review the numerous methods of phylogenetic inference. For this purpose, it was important to emphasize the key con-

cept — which is not always intuitively obvious — that overall similarity measures do not necessarily reflect the true evolutionary relationship of the taxa from which they were made. For the sake of explaining this concept (and realizing that I might anger some advocates of distance methods), I contrasted similarity and relationship in a simple manner, explaining the unique power of parsimony analysis to detect false similarities. Although I did not discuss other methods that attempt to build true phylogenetic trees from molecular data^{10,11}, such as maximum likelihood^{2,3}, invariant analysis, and distance methods using matrices adjusted for superimposed substitutions^{4,10,11} (Sidow's so-called "statistical" methods), I also did not advocate parsimony to the exclusion of these methods. Indeed, some types of molecular data, such as DNA hybridization and immunological distances, can be analysed only by distance methods¹⁰. I would again refer the interested reader to the excellent review articles^{10,12,13} mentioned in my review for a more thorough discussion of the various methods of phylogenetic inference.

I also highly recommend a recent article by Huelsenbeck and Hillis¹⁴, in which 16 distance, invariants and parsimony methods were compared regarding their abilities to find the correct four-taxon tree. (Maximum likelihood was not included in this study because of its computational expense.) This study is noteworthy because the 16 methods were extensively examined under several models of DNA evolution, including ones that conform to and ones that violate the assumptions of the methods. Parsimony methods were generally found to be good at inferring the true phylogeny, so long as the true phylogeny does not suffer from the following conditions: that of having a very short internal lineage separating two clusters, each cluster having one lineage that is quite long relative to its sister lineage. Because few phylogenetically informative events will occur along the short central lineage, homoplastic substitutions on the long lineages tend to group them erroneously during parsimony tree building⁶. This well-characterized phenomenon is termed the "Felsenstein zone"¹⁴ in phylogenetic inference. One important observation that Huelsenbeck and Hillis¹⁴ make is that all distance methods can fail to find the correct tree in the Felsenstein zone, and many find the wrong tree with certainty. We fully understand why parsimony can fail under these⁶ and some other conditions^{1,8}, but we are only beginning to understand exactly when and why distance methods can fail to find correct trees^{8,14}.

Another noteworthy point illustrated in ref. 14 is that with highly divergent sequences corrected distances become undefined; distance methods cannot be used in

such cases. Under less extreme conditions and models of evolution, the better distance and parsimony approaches were found to be about equally good at inferring the correct four-taxon phylogenetic tree¹⁴. A notable exception among the distance methods is UPGMA, which is consistently the worst. Given the renewed popularity of UPGMA, due to its use in many multiple sequence alignment programs, this should be an alarming result to molecular biologists. Although there is much to be learned from ref. 14, one should bear in mind that the four-taxon test case may not generalize well to trees with more taxa and more structure¹⁰.

Sidow makes the valid point that consideration should be made concerning the mode of evolution of the molecules under study, regardless of tree-building method used. However, he leaves the false impression that parsimony methods cannot consider such things as transition-transversion bias and different rates of evolution at different sites; in fact, such matters are simple to consider in PAUP¹⁵ and MacClade¹⁶ using stepmatrices and character weighting. Thus, modern parsimony methods can indeed incorporate models of DNA evolution quite similar to those used in "statistical" character-based methods such as maximum likelihood¹⁰.

Although parsimony does not correct the original data for superimposed evolutionary events before tree building (as corrected distance methods do), the very concept of hidden events in pairwise comparisons of sequences arose years ago through parsimonious reconstructions of amino acid replacements on trees and consideration of the genetic code. The concept is now applied directly to DNA sequences. Thus, parsimony and distance methods of sequence analysis should be viewed as complementary, not exclusionary. It is through their thoughtful combined use that we gain our greatest in-

sights about evolution at the molecular level.

Moreover, in contrast to other methods of tree construction, parsimony can be used to build trees from any type of character data^{15,16}; thus, it is widely used by evolutionary biologists of all stripes, not just molecular phylogeneticists. Perhaps most importantly, parsimony analysis is the logical method by which we reconstruct evolutionary history, regardless of how the phylogeny was inferred. Thus, unlike the methods that Sidow advocates, parsimony analysis has powers that extend far beyond the inference of gene trees¹.

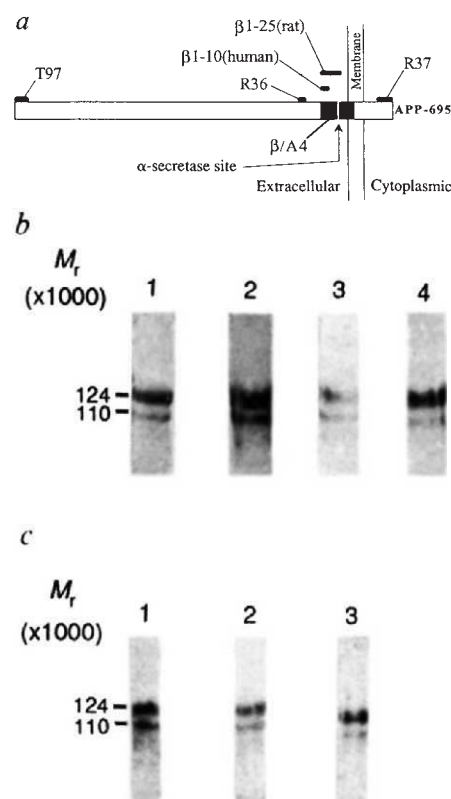
Caro-Beth Stewart

Department of Biological Sciences,
State University of New York,
Albany, New York 12222, USA

Gelatinase A not alpha-secretase?

SIR — The deposition in the brain of amyloid fibrils assembled from the β /A4 peptide is an early and seminal event in the pathogenesis of Alzheimer's disease¹. This peptide is thought to be derived by proteolysis from a membrane-spanning amyloid precursor (APP)² which exists as multiple isoforms produced by alternative messenger RNA splicing³. Soluble, secretory derivatives of APP have been detected in cerebrospinal fluid, brain, serum and in media conditioned by cultured cells³. Rat adrenal pheochromocytoma PC-12 cells secrete APP by proteolytic cleavage at the Lys 16–Leu 17 bond in the middle of the β /A4 sequence⁴. The enzyme responsible (APP α -secretase) remains uncharacterized, but Miyazaki *et al.* have now suggested⁵ that it may be the matrix metalloproteinase, gelatinase A. We have examined this possibility in two ways, first by testing a specific gelatinase inhibitor, tissue inhibitor of metalloproteinases-1 (TIMP-1), for its ability to interfere with the secretion of soluble forms of APP by PC-12 cells; and second, by assessing the activity of gelatinase A in media conditioned by PC-12 cells.

The cells were cultured and treated with 0, 2.5 or 5 μ g ml⁻¹ TIMP-1 as described in the figure legend. Secreted APP was detected in conditioned media by western blotting using several antibodies raised against appropriate synthetic peptides (*a* in the figure). Antibodies against T97, R36, β 1-25 and β 1-10 all revealed the characteristic pattern of two immunoreactive bands (*b* in the figure) of relative molecular mass 124,000 and 110,000. We obtained no reactivity with anti-R37, showing that both bands are presumably C-terminally truncated APP containing at least part of the amyloid sequence. The



TIMP-1 has no effect on the release of secretory APP into the culture medium of PC-12 cells. *a*, Specificity of the antibodies used; *b*, immunoreactive bands observed (1, anti-T97; 2, anti-R36; 3, anti- β 1-25; 4, anti- β 1-10); *c*, lack of effect of TIMP-1 (1, no TIMP; 2, 2.5 μ g ml⁻¹ TIMP; 3, 5 μ g ml⁻¹ TIMP).

METHODS, PC-12 cells were cultured in RPMI medium supplemented with 5% fetal calf serum and 10% heat-inactivated horse serum. Cells were grown at 37 °C under 5% CO₂ in Falcon 3028 flasks coated with poly-L-lysine. When confluency had been attained, cells were treated with 50 ng ml⁻¹ nerve growth factor (NGF) and incubated for 4 days, after which the medium and NGF were renewed. After a further 3 days, cells were washed with serum-free medium and incubated with RPMI medium supplemented with NGF and 2.5% Ultrosor (Gibco, serum replacement). Conditioned media were dialysed and then passed down a heparin column. Secreted APP was eluted with 1 M NaCl, dialysed, concentrated and examined by western blotting.

presence of these bands was not altered by treatment with TIMP-1 (*c* in the figure).

We assessed gelatinase activity by ¹⁴C-labelled gelatin assay of the conditioned medium from the PC-12 cell cultures described above, and by gelatin zymography⁶. The latter method can detect as little as 1 pg enzyme and is not affected by the presence of endogenous inhibitors. We detected no activity using either technique.

In the light of these results it seems unlikely that gelatinase A is the α -secretase in PC-12 cells. Nevertheless, these results do not exclude the possibility

1. Stewart, C.-B. *Nature* **361**, 603–607 (1993).
2. Felsenstein, J. *J. molec. Evol.* **17**, 368–376 (1981).
3. Hasegawa, M. *et al. J. molec. Evol.* **22**, 160–174 (1985).
4. Jin, L. & Nei, M. *Molec. Biol. Evol.* **8**, 356–365 (1991).
5. Goldman, N. *Syst. Zool.* **39**, 345–361 (1990).
6. Felsenstein, J. *Syst. Zool.* **27**, 401–410 (1978).
7. Hendy, M. D. & Penny, D. *Syst. Zool.* **38**, 297–309 (1989).
8. DeBry, R. W. *Molec. Biol. Evol.* **9**, 537–551 (1992).
9. Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406–425 (1987).
10. Swofford, D. L. & Olsen, G. J. in *Molecular Systematics* (eds Hillis, D. M. & Moritz, C.) 411–501 (Sinauer, Sunderland, Massachusetts, 1990).
11. Felsenstein, J. *PHYLP (Phylogeny Inference Package), Version 3.5* (Computer software package and manual distributed by the author, Dept Genetics, University of Washington, Seattle, 1993).
12. Nei, M. in *Phylogenetic Analysis of DNA Sequences* (eds Miyamoto, M. M. & Cracraft, J.) 90–128 (Oxford University Press, New York, 1991).
13. Felsenstein, J. *A. Rev. Genet.* **22**, 521–565 (1988).
14. Huelsenbeck, J. P. & Hillis, D. M. *Syst. Biol.* **42**, 247–264 (1993).
15. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1* (Illinois Natural History Survey, Champaign, 1993).
16. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Evolution* (Sinauer, Sunderland, Massachusetts, 1992).

of an active intracellular or membrane-associated form of gelatinase that is not accessible to TIMP-1 in the culture medium.

Dominic M. Walsh

Carvell H. Williams

Hilary E. Kennedy

David Allsop*

*Division of Biochemistry,
The Queen's University of Belfast,
Belfast BT7 9BL, UK*

Gillian Murphy

*Strangeways Research Laboratory,
Cambridge CB1 4RN, UK*

* Also at: Molecular Neuropathology Department, SmithKline Beecham Pharmaceuticals, Harlow, Essex CM19 5AD, UK.

1. Hardy, J. & Allsop, D. *Trends pharmac. Sci.* **12**, 383–388 (1991).
2. Kang, J. *et al. Nature* **341**, 144–147 (1987).
3. Bush, A. I., Beyreuther, K. & Masters, C. L. *Pharmac. Ther.* **56**, 97–117 (1992).
4. Anderson, J. P. *et al. Neurosci. Lett.* **128**, 126–128 (1991).
5. Miyazaki, K., Hasegawa, M., Funahashi, K. & Umeda, M. *Nature* **362**, 839–841 (1993).
6. Ward, R. V., Hembry, R. M., Reynolds, J. J. & Murphy, G. *Biochem. J.* **278**, 179–187 (1991).

NO comments

SIR — Lipton *et al.*¹ mention two redox-based forms of nitric oxide, one neuroprotective (NO^+ , nitrosonium ion) and the other neurotoxic ($\text{NO}^\bullet$; neutral nitric oxide with a single electron in an antibonding orbital). But there is another redox form of $\text{NO}^\bullet$, the nitroxyl anion (NO^-) which Lipton *et al.* do not mention in their *Nature* paper. In an earlier paper², however, some of the same authors reported that the nitroxyl anion rapidly converts to nitrous oxide (N_2O) through dimerization and dehydration. They showed the importance of this process for biological systems by demonstrating that $\text{NO}^\bullet$ converts to NO^- by enzymatic reductase pathways in bacteria.

Bacterial nitric oxide reductases can convert nitric oxide to nitrous oxide as part of the nitrogen cycle³. If these nitric oxide reductases have been conserved in mammalian species, then the possible role of nitrous oxide as an endogenous neurotransmitter would have to be evaluated because exogenous nitrous oxide has previously been shown to have a direct influence on neurotransmission at biologically relevant concentrations^{4,5}. The conversion of nitric to nitrous oxide would also offer another means of detoxifying nitric oxide, because nitrous oxide is orders of magnitude less toxic than nitric oxide⁶.

M. A. Gillman

F. J. Lichtigfeld

*South African Brain Research Institute,
Waverley 2090,
Johannesburg, South Africa*

SIR — Lipton *et al.*¹ address the observation that sometimes $\text{NO}^\bullet$ is toxic, and sometimes it isn't. They argue convincingly

that toxicity comes from $\text{NO}^\bullet$ reacting with superoxide to form ONOO^- , which is a strongly oxidizing agent⁷ involved in deleterious processes⁸, and that protection results if a compound in which $\text{NO}^\bullet$ is present as NO^+ nitrosylates a sulphhydryl group of the NMDA receptor, which indirectly prevents $\text{NO}^\bullet$ formation. Thus, $\text{NO}^\bullet$ is bad, and NO^+ is good. Triplet NO^- should be added to the list and classified as bad, because it reacts rapidly with dioxygen to form ONOO^- (ref. 9).

In his helpful News and Views article¹⁰ on the paper by Lipton *et al.*¹, Snyder suggests that "the designation nitric oxide should be restricted to the reduced, $\text{NO}^\bullet$ form of the molecule, while the parent NO should be called 'nitrogen monoxide', and the oxidized form NO^+ , the nitrosonium ion." As a member of the IUPAC committee for the nomenclature of inorganic chemistry I would like to caution against these recommendations for two reasons.

First, the difference between ' $\text{NO}^\bullet$ ' and 'parent NO ' is unclear. Second, there are recommended names for NO^+ , $\text{NO}^\bullet$ and NO^- , namely nitrosyl cation, nitrogen monoxide and oxonitrate(1^-) anion, respectively¹¹. We do not recommend the commonly used name for $\text{NO}^\bullet$, nitric oxide, because the suffix '-ic' can suggest an oxidation state that is not always the same. As an example, the oxidation state of nitrogen in nitric oxide is 2+, but it is 5+ in nitric acid.

Nitrogen monoxide is a perfectly systematic name for $\text{NO}^\bullet$ when one considers that $\text{NO}_2^\bullet$ is named nitrogen dioxide. The name oxonitrate(1^-) for NO^- sounds odd, but it makes sense next to the systematic names dioxonitrate(1^-) for nitrite, and trioxonitrate(1^-) for nitrate. Similarly, the recommended name for peroxyxynitrite or peroxyxynitrite is oxoperoxyxynitrate(1^-). I realize that many working in the rapidly expanding nitric oxide — oops! — nitrogen monoxide field will be reluctant to accept and use these names, but they are easier to learn than the old ones.

W. H. Koppenol

*Biodynamics Institute and Departments of
Chemistry and Biochemistry,
Louisiana State University,
Baton Rouge, Louisiana 70803, USA*

LIPTON *ET AL.* REPLY — The divergence in the views expressed by Gillman and Lichtigfeld and by Koppenol on the 'virtue' of NO^- illustrates an important point: the propensity of various NO congeners for toxic or protective actions is determined by the chemistry that they undergo in specific cellular milieux². We also suggest that the thiol mediated formation of hydroxylamine (NH_2OH) from NO^- could, in many biological systems, be the preferred NO^- reaction pathway, and thus serve to detoxify $\text{NO}^\bullet$. NO^+ (or the group transfer there of) is mutagenic when N -

nitrosation of DNA leads to deamination, but is protective when S -nitrosation leads to down-regulation of NMDA receptor activity. $\text{NO}^\bullet$ exerts numerous salutary effects through activation of guanylate cyclase, but also manifests peroxyxynitrite-mediated cytotoxic actions. Designation of the various redox-related forms of NO as 'good' or 'bad' should in our view be avoided.

Koppenol brings the authority of IUPAC, and an appealing whimsy, to bear on the issue of Nomenclature. We use ' NO ' as a generic, family name that embraces the specific species NO^- , $\text{NO}^\bullet$ and NO^+ . This usage, carefully defined in our publications (see ref. 1), is handy in broad discussions of NO -group transfer reactions — particularly in biological systems where the nature of the species transferred (NO^- , $\text{NO}^\bullet$, NO^+) is not known with certainty, and the character of the NO moiety in bioactive substances is ambiguous². We deliberately introduced this usage¹ to expand the perspectives on NO action: the integration of the distinctive chemistries of NO^- , $\text{NO}^\bullet$ and NO^+ within the field of 'nitric oxide' biology is critical to an understanding of the 'Janus faces' of NO .

We endorse the suggestion of Snyder¹⁰ that biomedical researchers recognize these distinctions and adhere to an unambiguous Nomenclature. We concede that a defect of our scheme is the possible confusion that could arise from the common tendency to write $\text{NO}^\bullet$ simply as ' NO ' — IUPAC recommendations notwithstanding. With the conviction that people in the field have come to appreciate the significance of the broader chemistry of ' NO ', we defer to IUPAC — despite misgivings about OONO^- being termed 'oxoperoxyxynitrate(1^-)' (personally we prefer 'Oh NO ') — and formally offer NO further comment on the issue of Nomenclature.

S. A. Lipton

J. S. Stamler

D. J. Singel

*Departments of Neurology, Chemistry and
Medicine,
Harvard Medical School,
Boston, Massachusetts 02115, USA*

1. Lipton, S. A. *et al. Nature* **364**, 626–632 (1993).
2. Stamler, J. S., Singel, D. J. & Loscalzo, J. *Science* **258**, 1898–1902 (1992).
3. Ferguson, S. J. in *The Nitrogen and Sulphur Cycles* (eds Cole J. A. & Ferguson S. J.) 1–29 (Cambridge Univ. Press, 1988).
4. Daras, C., Cantrill R. C. & Gillman, M. A. *Eur. J. Pharmac.* **89**, 177–178 (1983).
5. Ori, C., Ford-Rice, F. & London, E. *Anesthesiology* **70**, 541–544 (1989).
6. Marshall, B. E. & Longnecker, D. E. in *The Pharmacological Basis of Therapeutics* 8th edn (eds Goodman, *et al.*) 285–310 (Pergamon, New York, 1990).
7. Koppenol, W. H. *et al. Chem. Res. Toxic.* **5**, 834–842 (1992).
8. Beckman, J. S. *et al. Nature* **364**, 584 (1993).
9. Donald, C. E. *et al. Inorg. Chem.* **25**, 2676 (1986).
10. Snyder, S. H. *Nature* **364**, 577 (1993).
11. Leigh, G. J. (ed.) *Nomenclature of Inorganic Chemistry* (Blackwell Scientific, Oxford, 1990).

Remember me

Jane Maienschein

If I Am To Be Remembered: The Life and Work of Julian Huxley with Selected Correspondence. By K. R. Dronamraju. *World Scientific*: 1993. Pp. 294. £27, \$38.
Julian Huxley: Biologist and Statesman of Science. Edited by C. Kenneth Waters and Albert Van Helden. *Rice University Press*: 1993. Pp. 344. \$32.50.

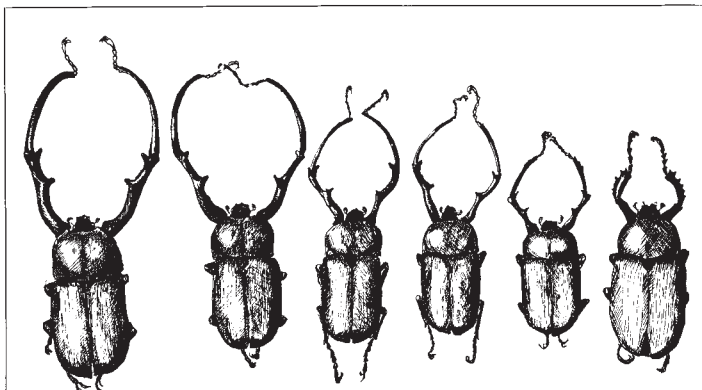
If Krishna Dronamraju's *If I am to be Remembered* had appeared earlier, we might have been able to say "Well, it's about time that someone took Julian Huxley seriously". Unfortunately for Dronamraju, Waters and van Helden's excellent edited volume appeared at about the same time and has made Dronamraju's much lesser contribution superfluous.

Huxley wrote in his *Memories* (1970): "If I am to be remembered, I hope it will not be primarily for my specialized scientific work, but as a generalist." Dronamraju quotes Huxley on the first page of his introduction, and then goes on to say that the primary scope of his book is Huxley's biological work. Alas, this is typical of his volume. The focus is not clear—Dronamraju actually offers relatively little about this biological science—and the discussion is not very rigorous. Instead he offers a somewhat arbitrary collection of prefaces, a weak, repetitious biographical sketch that lacks many clear dates or any sense of context, and a rather jumbled selection of letters from the Huxley Archives at Rice University. In short, this is a disappointing offering. It is especially disappointing because Dronamraju evidently attended the conference at Rice that Waters and van Helden organized and that forms the basis for their work. Dronamraju should thus have known that a more serious scholarly historical study on Huxley was in preparation, but he does not mention the conference or its resulting book.

That book, *Julian Huxley: Biologist and Statesman of Science*, inspires some of the normal reactions to edited efforts: it is uneven, some of the short commentaries add little and could better have been omitted, whereas others are nice additions; a few of the papers seem to have been written for some other purpose and give a nod to Huxley only as an afterthought. Yet, overall, the volume provides a useful and persuasive introduction to this twentieth-century biologist who identified the "modern synthesis" in evo-

lutionary biology and who had a hand in many other ventures both public and private. Along the way, some of the contributions are first rate and challenge us to explore further, enticing us to study the rich archival collection kept at Rice University.

What draws the volume together as a whole is Waters' perceptive introductory essay. He provides a biographical sketch that frames the other studies and allows them to probe more substantively into the content of Huxley's historical, biological



BEETLEMANIA — heterogeny of the male forelimb of the beetle *Euchirus longimanus*. The specimen on the right is a female, the rest are males of increasing absolute size. From *Problems of Relative Growth* by Julian Huxley, first published in 1932 and now reprinted. Johns Hopkins University Press, £16.50.

and public contributions. This excellent overview presents an intriguing portrait of Huxley as a real person, seized by periods of self-doubt alternating with episodes of prodigious enterprise. Huxley emerges neither as the glorious architect of the synthesis, nor as a deflated and somehow disappointingly deficient member of the great Huxley clan. This volume as a whole invites us to look at Huxley — and twentieth-century science along with its social context — in a different and richer way.

Strong essays by Colin Divall and Robert Olby place Huxley in his society, as a humanist who may have been a Victorian thinker trapped in an "unsympathetic modern age" (Divall) or as a complex outsider of whom his wife Juliette said that "[t]here were, in fact, so many Julians within this tall carelessly dressed wanderer bent on his various ploys, escaping from one activity by diving into yet

another" (Olby).

Jan Witkowski, Frederick Churchill and Richard Burkhardt provide stimulating analyses of Huxley's considerable biological contributions. Witkowski explains Huxley's embryological and cell study in a way that makes it seem perfectly reasonable that an intelligent Huxley of the early twentieth century should be attracted to such research problems as the study of cell (de)differentiation and relative growth. Where Dronamraju asserts that some of this work was Huxley's best but does not explain it very well or make it clear just what the contribution was, Witkowski does both. Churchill's perceptive essay furthers the discussions by showing how Huxley sought to combine the organismic gradient interpretation of Charles Manning Child with Hans Spemann's somewhat antithetical organizer theory. Churchill argues that Huxley embraced the "English emphasis on the organism as a dynamic, biochemically integrated whole rather than on the cell as the basic unit of life." Burkhardt then considers another phase of Huxley's biological work, his field study of the natural history of bird behaviour. A total of only 40–50 days (or even partial days) in the field produced useful ethological contributions but not, as Burkhardt argues persuasively, the major synthetic or pioneering work that Huxley later saw himself as having provided.

G. Allen, D. Paul and D. Kevles examine Huxley's popular work in a section entitled "Huxley the Statesman of Science." Allen and Paul interpret Huxley's eugenic views somewhat

differently, and the two together offer a rich sense of the reasoning and (implicitly) the role of secular humanism behind this prominent eugenicist's public views.

This volume as a whole provides an introduction to this particular Huxley. Although there is little detailed discussion of his administrative roles at the London Zoo or UNESCO, for example, or of the effects on his science of his nervous breakdowns and insecurities, the collection offers points of contact for such study. And the Rice archives offer a rich starting point, as Nancy Boothe's appendix explains. With Huxley's own two-volume autobiography, the foundation is well laid for the "further research" to which the editors invite readers. □

Jane Maienschein is in the Departments of Philosophy and Zoology, Arizona State University, Box 872004, Tempe, Arizona 85287-2004, USA.

Accidents will happen

David L. Sills

The Limits of Safety: Organization, Accidents, and Nuclear Weapons. By Scott D. Sagan. Princeton University Press: 1993. Pp. 286. \$29.95.

THREE topics of current concern and interest are analysed in *The Limits of Safety*: the dark side of technology; the theory of accidents; and conditions for the prevention of nuclear war.

The core of the book consists of accounts of near-catastrophic accidents in the US nuclear weapons system: planes armed with nuclear bombs have inadvertently entered Soviet air space; nuclear weapons have threatened US personnel in various ways; and incorrect reports of imminent Soviet attacks have almost triggered US retaliatory strikes. The sources used in the book include the scholarly literature on accidents and near-accidents with US nuclear weapons, the huge literature specifically on the 1962 Cuban missile crisis, dozens of interviews with participants in these events and large numbers of formerly classified US government documents. A substantial portion of the text is given over to a detailed and technical analysis of these accidents and near-accidents. The author, a political scientist at Stanford University, has published extensively on US nuclear weapons policy.

The past decade has brought a plethora of books that question or attack the benefits of technology and that describe or deny its dark or underside. Some of these books are popular, others scholarly; some describe the threats of automobiles, nuclear fission, pesticides, tobacco or whatever — threats to our safety and well-being. Others attack the critics of technology as being latter-day Luddites. Controversies abound.

Two characteristics of this book are important and refreshing. First, the style is coolly nonpolemical; Sagan is deeply worried about the prospects for nuclear weapons accidents, but he avoids the language of gloom-and-doom. Second, not unrelated to the first, the author's orientation is towards the appropriateness of various theories of command-and-control measures rather than towards the incompetence of individual participants or the mechanical flaws of weapons themselves. Sagan approvingly quotes the remark of the Yale sociologist Charles Perrow that "the dangerous accidents lie in the system, not in the components", thus effectively distancing them both from the charge of Luddism.

Addressing the theory of accidents, Sagan started his research with a puzzle:

how is it — in a world that has witnessed the nuclear accident at Chernobyl, the Exxon Valdez oil spill, the Love canal chemical leak, the explosion of the space shuttle *Challenger*, and the toxic gas disaster at Bhopal — that "there has never been an accidental or unauthorized detonation of a nuclear weapon, much less escalation to an accidental nuclear war". To seek answers to this puzzle, Sagan examined two competing schools of thought about accidents: high reliability theory and normal accidents theory. It is necessary to describe these theories briefly.

High reliability theory is based upon the optimistic view that "extremely safe operations are possible, even with extremely hazardous technologies, if appropriate organizational design and management techniques are followed". Four critical elements of high reliability theory are

visible. Third, when redundancy makes a system appear more safe, operators and their bosses often take advantage. "Fixes, including safety devices", Sagan quotes from Perrow's *Normal Accidents* (1984), "often merely allow those in charge to run the system faster, or in worse weather, or with bigger explosives."

Normal Accidents contains the first full formulation of Perrow's normal accidents theory, a theory that concentrates on the properties of technological systems themselves rather than on the errors that owners, designers and operators make in running them. In part, *The Limits of Safety* is a test of Perrow's more general theory by applying it to nuclear weapons systems, but it importantly adds a political dimension to the theory, demonstration that organizations create structures that themselves contribute to the likelihood of accidents.

The goal of normal accidents theory is to find a more basic explanation for accidents than operator error or faulty equipment. Primarily, it is a search for details of the ways in which the components of a system are related: whether there is tight or loose coupling between components and whether there is linear or complex interaction between them: tight coupling and complex interaction make accidents more likely. Viewed from these perspectives, most accidents are rather pessimistically found to be normal, in the sense that no particularly unusual events caused them.

Sagan finds that Perrow's pessimistic conclusion — that "serious accidents are inevitable, no matter how hard we try to avoid them" (p.31) — contrasts sharply with the optimism of the high reliability theorists. He quotes with approval what is perhaps Perrow's most succinct formulation: "Accidents are inevitable and happen all the time, serious ones are inevitable though infrequent; catastrophes are inevitable but extremely rare." Sagan thus believes that a serious or catastrophic accident in the US nuclear weapons system is inevitable. (He obviously believes this to be true of other nuclear weapons systems as well, but he has no comparable data from their files.)

Perrow's analysis of tightly coupled, complex technologies led him to conclude that nuclear weapons can never be accident-free and thus the only rational course of action is to abandon them. This is also the policy recommendation of such writers as Jonathan Schell and Telford B. Taylor, among hundreds of others. Interestingly, Sagan reaches the opposite conclusion: abandoning the nuclear weapons arsenal, he believes might increase the likelihood of nuclear war.

His argument is both technical and political. Technically, the fact that nuclear weapons are both relatively small and difficult to detect means that some nations

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Nuclear weapons can never be accident-free.

identified: (1) giving a high priority to safety and reliability as organizational goals; (2) creating extensive redundancy in the number of both personnel and technical safety measures; (3) developing a "high reliability culture"; and (4) practising sophisticated forms of trial-and-error organizational learning.

Sagan gives many examples of all four elements but gives most emphasis to the limits of redundancy in preventing accidents. He believes that adding redundant parts or safety devices to a complex technological system can actually increase the likelihood of accidents. First, redundant systems are often less independent than their designers believe and thus they usually increase interactive complexity among the system's components. Second, adding redundancy often makes a system more opaque and its components less

will violate any attempt to prohibit them. Politically, sovereign nuclear powers will have strong incentives to maintain at least some nuclear weapons. And even if all nuclear weapons were destroyed, "the knowledge of how to build them would remain in the world [and] in any conventional conflict, states would have very strong incentives to build nuclear weapons, and the first state to get them would have incentives to use them before other states acquired them". Sagan concludes that it is "not even clear whether the likelihood of nuclear war would be increased or decreased if current policies of nuclear deterrence were replaced by complete disarmament".

What can be done about this threat? Because it is not the power of nuclear weapons that creates the likelihood of accidents but the structure of the nuclear arsenal, it is the structure that must be changed.

Sagan's suggestions for changing the structure are necessarily general. He believes that the dangers inherent in interac-

tive complexity can be reduced by physically separating components, for example, national warning systems should be located in different localities from nuclear warheads. Tightly coupled systems require time to recover from individual failures, and Sagan proposes increasing the time required for a launch by removing all warheads from deployed missiles. And as accidental launches may still occur, Sagan proposes installing radio-controlled devices on missiles that could destroy the missiles in flight before they reach their targets.

The content of *The Limits of Safety* thus ranges from the general theory of accidents to how-to-do-it suggestions for any nation's nuclear planners. It is a skilful blending of social science, physical science, organization science and military science and is highly recommended to readers in all four fields. □

David L. Sills, a sociologist, is at 14 Crockett Street, Rowayton, Connecticut 06853, USA.

Credibility as a science

Stuart Sutherland

The Story of Psychology. By Morton Hunt. Doubleday: 1993. Pp. 762. \$29.95.

SINCE the beginning of this century, experimental psychology has aped the manners of other scientific disciplines. It has, for example, been influenced by electromagnetic field theory (the Gestaltists), logical positivism (the behaviourists), Chomskian linguistics (the psycholinguists), and cybernetics, information theory and the digital computer (cognitive psychologists). In addition, in areas such as vision and hearing there are close connections with neurophysiology, although here the influence has been in both directions with psychologists often anticipating the physiologists' findings.

The emulation of other disciplines has occurred largely in an attempt to give psychology credibility as a science, but it is likely to continue to differ from other sciences in two ways. First, different tasks must be executed by different processes so each requires a different theoretical explanation: there can be no such general theories as quantum mechanics (even connectionism requires different connectivity for different tasks). Second, for much of human behaviour it is impossible to take into account the effects of different environments (including different cultures) and the complex interaction between nature and nurture.

Nevertheless, we do now have reasonably rigorous theories of some aspects of perception, such as colour vision or the

judgement of the direction of a sound. Moreover, even where rigour is not possible, remarkable new insights supported by experimental evidence have sometimes been obtained, for example into the many irrelevant factors that influence decision making. When, however, one turns to such areas as personality or emotion, it is doubtful whether any advance has been made and certainly no agreement has been reached. As Morton Hunt argues, psychology is a collection of discrete and very different disciplines.

In *The Story of Psychology*, Hunt attempts to review the history of all of them from Thales and Democritus to the present day (when the subject has grown to such an extent that one in six doctoral degrees in the United States are in psychology). He accomplishes his monumental task by omitting everything that might be difficult for the reader to understand; there is, for example, nothing on hearing, the horopter, or modern research on animal learning. The book is written as a series of essays, some on an individual psychologist, some on a topic. They are somewhat disconnected: for example, there is no mention of the relationship between J.

Wolpe's behaviour therapy and B. F. Skinner's later advocacy of behaviour modification. For each field of psychology, Hunt takes a few instances of theories and experiments that give the layman a reasonable impression of what it is all about. He writes clearly but without wit or charisma. Naturally he has lapses: his explanation of connectionism is unintelligible and he does not describe its remarkable success in accounting for phenomena that it was not designed to explain. Sometimes he includes very trivial work, such as the 'New Look' of the 1950s, in which efforts were made to prove that motivation influenced the apparent size of objects — a piece of food was supposed to look bigger to a hungry child than to a satiated one, a phenomenon lacking interest even if it were true, which many doubt. As befits an outsider, he refrains from judging the worth of the research he describes; instead, he assesses its value by counting heads, only too often dunderheads.

One of the best features of the book is that, wherever possible, Hunt lets the members of his huge cast speak for themselves through the use of quotations. I particularly liked the salutary advice of the Royal Society given soon after its inauguration, urging its members to communicate in "the language of artisans, countrymen and merchants [rather than] that of wits and scholars".

Nowadays, possibly as a result of watching television, the lay reader's apprehension scan is so low that he cannot read straightforward accounts of science. In consequence, many recent popular scientific books have pandered to him by providing a banal mixture of science and gossip. Hunt is no exception. There is too much of "On a sunny morning in Ancient



NO RELATION — seeds and female sporophylls of *Cycas taiwaniana*. Cycads are a unique group of plants, totally unrelated to any other living plant, although often wrongly linked to palms and ferns. From *Cycads of the World: Ancient Plants in Today's Landscape* by David L. Jones. Smithsonian Institution Press, \$45.

Athens a sober, bearded Aristotle encountered his master, a grinning clean-shaven Plato in the agora", or to take an actual example, "A short, round ebullient man Dunlap . . .".

Perhaps the most interesting chapter is that on recent developments in applied psychology, many of which are valuable. It has, for example, been shown that people with their own office (known as "private space" in the jargon) both work more effectively and obtain greater job satisfaction than those in an open-plan space. Many large companies and most newspapers are either unaware of this finding or prefer to ignore it. A large cause of the Three Mile Island accident was that the displays had been designed by engineers not psychologists. Thirty per cent of them were too high for the operators to read, and the same colour coding was used to represent normal functioning on some displays and abnormal functioning on others. Companies, and even universities, continue to use the interview as a method of selection, despite the fact that it is useless if more objective data are available. Although good applied psychology can be extremely helpful, there is a debit side. Don't believe anyone who claims that biofeedback works, that you can learn in your sleep, that subliminal stimuli can influence motivation, or that half-baked seminars given by a so-called psychologist can improve creativity or productivity in a company. All such techniques, however plausible, have to be validated before they can be known to be of use. If you want to know whether the introduction of any change is helpful, only the psychologist is equipped to find out. It sounds easy, but it is no mean feat. Surprisingly, Hunt fails to point out the most important achievement of 3,000 years of psychology, namely, the development of good experimental techniques. Nowadays, psychology's methodology is even influencing other disciplines, such as medicine and biology, thus repaying some of the debt owed. □

Stuart Sutherland is in the Department of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

New in paperback

■ *Niels Bohr's Times, in Physics, Philosophy, and Polity* by Abraham Pais, Oxford University Press, £12.95, \$17.95. For a review see *Nature* **353**, 511 (1991).

■ *The New Turing Omnibus: 66 Excursions in Computer Science* by A. K. Dewdney, Freeman, £19.95, \$24.95. For a review see *Nature* **341**, 496 (1989).

Just the facts

Manfred Schliwa

The Adhesion Molecule FactsBook. By Rod Pigott and Christine Power. Academic Press: 1993. Pp. 190. £19.50, \$42.

Guidebook to the Cytoskeletal and Motor Proteins. Edited by Thomas Kreis and Ronald Vale. Oxford University Press: 1993. Pp. 276. £22.50, \$37 (pbk); £45, \$75 (hbk).

Guidebook to the Extracellular Matrix and Adhesion Proteins. Edited by Thomas Kreis and Ronald Vale. Oxford University Press: 1993. Pp. 176. £18.50, \$30 (pbk); £40, \$65 (hbk).

PICTURE this: you are attending a meeting on structural proteins. In the discussion after one of the talks the speaker reassures a discussant that 'his' protein has features reminiscent of ABC123^{xyz}, a component of periparaprosynfitin. You are alarmed: did you miss something important there? ABC123^{xyz} sounds vaguely familiar, but what is its relationship to periparaprosynfitin (or was it . . . synphitin?). You decide to start a computer search once you return to your laboratory and wish for a ready source that would refresh your memory on these proteins.

Now a remedy is here — at least as far as cytoskeletal, adhesion and extracellular matrix proteins are concerned. These books provide concise descriptions of most of the proteins that make up these three classes of structural proteins. Between 1 and 5 pages are devoted to a terse characterization of the essential features of each protein. Indeed, 'essential' is the *spiritus rector* of these publications. Descriptions of molecules dominate; function in a cellular context is only touched upon or is left to the introductory chapters. These are the kinds of books where one would expect to be able to read up, in 5–10 minutes, on ABC123^{xyz} (if it existed) and its interaction with periparaprosynfitin (*sic*). The style of presentation is similar in the two types of publications. The *FactsBook* provides a more rigid framework for the delivery of the facts and includes amino-acid sequences and, wherever possible, a diagram of the domain structure of each protein discussed. The *Guidebooks* rarely present sequence information but frequently include original micrographs of the molecules and their distribution in the cell. A combination of both these features may seem desirable for future editions.

These are no books for beginners, but they will be welcomed by those new to these fields and, I dare say, virtually everyone already working in them. Not many of us are able to keep a mental record of the features of all the

cytoskeletal/adhesion/matrix proteins and their aliases (for example, CD44 = H-CAM = G90^{HERMES} = Pgp-1 = ECMR III = In(lu)-related p80 = HUTCH-1 = LY-24 = p85). New proteins and isoforms are being discovered, sequenced and characterized at an astonishing rate, and therefore publications of an encyclopaedic nature are becoming increasingly popular and useful. Of course they will not replace textbooks, monographs or proceedings, but they will supplement them.

This is a new type of publication for the cell biology field, and naturally there are some problems to be overcome. A major difficulty recognized by the authors/editors is that there are no rigorous, uniform criteria for the selection of entries. Pigott and Power suggest such a criterion for adhesive molecules (expression in a cell confers an adhesive phenotype that can be inhibited by specific antibodies) but also realize that this criterion is met by only a minority of proteins presumed to be adhesive. Kreis and Vale intended to include only the well characterized structural proteins that had been purified, sequenced and had antibodies raised against them, but several of their entries do not fulfil these criteria.

A second, related problem concerns the uniform weighting of entries: proteins such as actin, tubulin, tropomyosin or MAP2 on which hundreds or thousands of papers have been published are allowed just as much space as, for example, tenuin, on which one publication (without sequence) is available, or pericentrin, for which no full paper has yet been published. Perhaps some of the new, more exotic fruits should be given more time to ripen before they are served to a wider audience.

A last concern is timeliness. Compilations of an encyclopaedic nature such as these will remain valuable only if updated. New information is accumulating rapidly. For example, the entire myosin section of the *Guidebook* may require a revision soon to include a discussion of the many new members of this rapidly growing protein family. The inside front cover of the *Guidebook* announces a computer database, for updates only, of the entries in the *Guidebook* which authors of papers are requested to provide. This is laudable, but it may be dangerous to leave the entry of updates to the discretion and/or leisure of the authors. The hardware (books) should be updated, too, perhaps by converting them to loose-leaf editions where updates replace older entries and new proteins can be added. For these publications are bound to remain in demand and are likely to become even more effective and valuable in the future. □

Manfred Schliwa is at the Institute for Cell Biology, University of Munich, Schillerstrasse 42, 80336 Munich, Germany.

Impacts on the Earth by asteroids and comets: assessing the hazard

Clark R. Chapman & David Morrison

There is a 1-in-10,000 chance that a large (~2-km diameter) asteroid or comet will collide with the Earth during the next century, disrupting the ecosphere and killing a large fraction of the world's population. Although impacts of this magnitude are so infrequent as to be beyond our personal experience, the long-term statistical hazard is comparable to that of many other, more familiar natural disasters, raising the question of whether mitigation measures should be considered.

THE remnants of the formation of planets by planetesimal accretion¹ consist of numerous small objects, called asteroids and comets, which occasionally are perturbed into paths that cross the orbits of the Earth and other planets. Spacecraft exploration of the planets has revealed crater-scarred surfaces that testify to a rain of projectiles that continues today; the initial deluge during planetary accretion ended $\sim 3.8 \times 10^9$ yr ago with the Late Heavy Bombardment, but a shower of impacts has continued at an approximately steady rate ever since². Despite rapid rates of erosion and tectonism on Earth, over 140 terrestrial impact scars have been identified³. Evidence that the Earth still resides in a swarm of asteroids has grown due to recent improvements in telescopic search techniques; dozens of Earth-crossing asteroids are being found each year⁴. As is statistically inevitable, the ever-increasing discovery rate results in ever more frequent news reports about "near misses"⁵.

The effect of impacts on the Earth's geological history, its ecosphere and the evolution of life has become a major topic of current interdisciplinary interest, since publication of the Alvarez *et al.* idea⁶ that the Cretaceous/Tertiary (K/T) mass extinction was caused by the impact of an asteroid or comet ~ 10 km across. The Alvarez hypothesis has become widely accepted⁷ since identification of the probable primary impact site in the Yucatan⁸. Impacts have even been proposed as the dominant trigger for other "punctuated equilibrium" changes in the evolution of species⁹. Also, giant impacts in the early history of our planet are now invoked for the impact frustration of the origin of life (the idea that the Late Heavy Bombardment on Earth prevented a long enough period of environmental tranquillity for life to gain a foothold)^{10,11} and the origin of the Moon by a Mars-sized object striking the Earth¹².

Concerns about a modern-day hazard from comets striking the Earth were voiced as early as 1705 by Edmund Halley in his *A Synopsis of the Astronomy of Comets*. Soon after the first Earth-crossing asteroids were discovered about 60 years ago¹³, roughly correct estimates of asteroid impact frequencies and consequences were published^{14,15}. In 1947 ~ 100 craters larger than 1 m across were formed by the Sikhote-Alin (Siberia) fall of iron meteorites, an event that revived studies of the 1908 airburst over the Tunguska River region of Siberia¹⁶; the conclusion was that the resulting Tunguska devastation of $>1,000$ km² was caused by an atmospheric impact of a comet or asteroid. In 1981, a NASA sponsored "Spacewatch Workshop"¹⁷ called attention to the contemporary hazard associated with such cosmic impacts, and first proposed that projectiles as small as 1 km in diameter might potentially destabilize the global ecosystem, thereby threatening the continuance of human civilization—an event with consequences far exceeding the direct damage from the impact itself.

Here we review and evaluate the contemporary impact hazard associated with cosmic projectiles of various sizes and types, and

we compare it to other natural and human-generated hazards. Most projectiles <50 m in diameter, with energies $\lesssim 10$ megatons ($1 \text{ MT} = 4.2 \times 10^{15} \text{ J}$), dissipate the energy harmlessly in the upper atmosphere. A larger object could do severe local damage and, if it struck an urban area without warning, cause many deaths; though such impacts occur somewhere on Earth every century or so (very rarely near cities as urban areas constitute a very small fraction of the Earth's area), we show below that they generally pose a much less serious threat than other natural disasters, such as floods and earthquakes.

A greater risk is from objects large enough to disturb the global ecosystem, precipitate general crop failures, and kill unprecedented numbers of people. The threshold size of object for onset of such global catastrophes is poorly known, but we will argue that it is probably in the range 0.5–5 km. Statistically, the risk from such large impacts is similar to risks from other natural and technological disasters. A qualitative difference, however, shared only by nuclear war, is that a global impact catastrophe could lead to the breakdown of civilization.

Impacts are an extreme case of a low-probability/high-consequence hazard. Individual impacts are now unpredictable as few of the threatening objects have been catalogued, although most could be (within two decades) by a proposed survey¹⁸. Studies of risk perception, which we briefly review, suggest that the public may become anxious about such a catastrophe as knowledge of Earth-approaching objects increases. The impact hazard is becoming an issue of considerable visibility and public debate¹⁹, especially because controversial technology (for example, the use of nuclear explosives in deep space) might be required to mitigate the hazard. We conclude by posing questions of public policy raised by the newly recognized importance of the impact hazard.

Impact flux estimates

To analyse the cosmic impact hazard, we must first determine the flux of comets and asteroids striking the Earth. Due to minimal erosion or geological activity since the end of widespread volcanism on the Moon 3×10^9 yr ago, the Moon's lava plains have recorded the integrated flux of crater-forming cosmic debris in near-Earth space²⁰. The average flux at the top of the Earth's atmosphere can be derived from the lunar flux; it compares well, within errors, with direct estimates from the Earth's own recent cratering record³ and with a census of existing Earth-crossing asteroids and comets^{21–23}.

The Earth-crossing asteroids include rocky and metallic objects derived from main-belt asteroids through collisional fragmentation and chaotic dynamics; others are probably extinct comet nuclei²⁴. Most Earth-crossers will eventually be ejected from the Solar System with the aid of Jupiter's gravity; nearly all the rest will strike a terrestrial planet (about one-third of those will strike the Earth²⁴). Their bulk properties²⁵ range from

metallic (like iron–nickel meteorites) to stony (like chondritic meteorites) to cometary (low-density silicates, organics and volatiles). The total range in bulk density is about a factor of 10 ($\sim 8 \text{ g cm}^{-3}$ for iron, down to $\sim 1 \text{ g cm}^{-3}$ for cometary ices); the range in strength is presumably even greater.

By the end of 1992, 163 Earth-crossing asteroids had been catalogued (E. Bowell, personal communication). The largest is 1627 Ivar, $\sim 8 \text{ km}$ in diameter. From analysis of observational sampling statistics, we know that the census is incomplete for objects smaller than Ivar²². The estimated degree of completeness for $\sim 1 \text{ km}$ objects is $< 5\%$, and for 100-m objects it is $< 0.1\%$ (ref. 18). In addition to asteroids, active comets strike our planet. Although the comet flux (estimated from comet discoveries and recoveries²⁶) is only a few percent of the asteroid flux for objects of the same size, comets hit faster than the $\sim 20 \text{ km s}^{-1}$ typical for asteroids (typically $30\text{--}40 \text{ km s}^{-1}$ for short-period comets, $50\text{--}60 \text{ km s}^{-1}$ for long-period comets)²³. Therefore, active comets have kinetic energies several times greater than for similar sized asteroids, and constitute a significant share ($\sim 25\%$) of the impact hazard^{18,23}.

For the present hazard discussion, we adopt the average total impact flux on Earth estimated by Shoemaker²¹ and shown in Fig. 1, which gives impact flux as a function of projectile kinetic energy ($\frac{1}{2}mv^2$); the equivalent diameter is also labelled, assuming a stony-density object striking at 20 km s^{-1} . This curve does not include a claimed enhancement in the current flux of small asteroids ($< 50 \text{ m}$) based on discoveries by the Spacewatch programme²⁷. But their flux may be overestimated because of an unrealistically low assumed albedo (T. Gehrels, personal communication), their encounter velocities are lower than for typical Earth-approachers²⁸, and such small objects do not penetrate the Earth's atmosphere (except the small minority composed of iron)²⁹. Figure 1 is consistent with Shoemaker's updated analyses^{22,30} and we believe it represents the impact frequencies of objects $> 0.5 \text{ km}$ to within a factor of 2, and of smaller objects capable of doing damage to within a factor of 5.

Because of stochastic variability in the process of asteroid and comet break-up, there is a chance for significant temporal variations in the impact flux. 'Comet showers' could lead to major short-term increases in the impact hazard, and it has been argued³¹ that even the K/T extinction was due to more than one near-simultaneous impact. For purposes of this review, however, it is adequate to treat terrestrial impacts as occurring randomly in space and time.

Nature of the hazard

The Earth experiences a constant barrage of cosmic debris. Small particles burn up as visible meteors high in the atmosphere, whereas occasionally rocks survive deceleration in the atmosphere and reach the ground as individual meteorites. There is an extremely small hazard associated with such meteorite falls, and no authenticated human fatality (although automobiles, with their larger cross-section, have been struck a few times, most recently by the Peekskill meteorite on 9 October 1992 (ref. 32)). Larger projectiles, however, can do severe damage, up to and including mass extinctions by objects many kilometres in diameter. In this section we discuss the nature of the hazard posed by projectiles as a function of their sizes and physical properties.

Fireballs and bolides. The Earth's atmosphere represents a significant barrier to cosmic projectiles. Even at megaton energies, most meteoroids break up and are consumed before they reach the lower atmosphere. They are called fireballs or, if they explode, bolides. Figure 1 indicates that a bolide with the energy of the Hiroshima nuclear bomb ($\sim 0.015 \text{ MT}$) occurs annually, whereas a few events of megaton energy are expected each century. We are generally unaware of these events, for most occur at very high altitudes and their shock waves do not reach the ground, although they have been observed from surveillance satellites³³.

TABLE 1 Thresholds for global catastrophe

	Energy (MT)	Asteroid diameter (km)	Comet diameter (km)	Typical interval (yr)
Lower limit	1.5×10^4	0.6	0.4	7×10^4
Nominal	2×10^5	1.5	1.0	5×10^5
Upper limit	10^7	5	3	6×10^6

Objects give up most of their kinetic energy in the atmosphere and explode if they encounter a column of atmosphere roughly greater than or equal to their mass (corresponding to $\leq 20 \text{ m}$ diameter for a 1 g cm^{-3} object). Actually, substantially larger meteoroids are blocked as a result of aerodynamic stresses that cause flattening, fragmentation, spreading of fragments and rapid braking at high altitude^{34,35}. The height of fragmentation depends primarily on the meteoroid's physical strength; only the strongest iron meteoroids reach the ground in one piece. For non-iron meteoroids, the minimum energy required to penetrate to the lower atmosphere is $\sim 10 \text{ MT}$, or $\sim 50 \text{ m}$ diameter for a stony object hitting at 20 km s^{-1} . Analogous calculations for the thicker atmosphere of Venus³⁶ were confirmed by the deficiency of smaller craters observed by the Magellan spacecraft.

Locally devastating impacts. If a large meteoroid is able to penetrate to within $\sim 25 \text{ km}$ of the surface (or actually strikes) at velocities of tens of kilometres per second, the resulting explosion can cause severe damage analogous to that from nuclear bomb explosions of similar energies, but without neutron or γ -radiation, or radioactive fallout. It is well known from civil defence studies that the area of devastation scales approximately as the explosive yield to the $2/3$ power, and is somewhat greater for a (low) airburst than for a groundburst explosion³⁷.

The 1908 Tunguska airburst provides a calibration, with a shock wave sufficient to fell trees over an area of $\geq 1,000 \text{ km}^2$ and a fireball that ignited fires over a smaller area near ground zero. The yield of the Tunguska blast has been estimated at $10\text{--}20 \text{ MT}$ from microbarograph measurements in Europe and other methods³⁴. If we assume that the radius of forest devastation would apply also to destruction of nearly all buildings (chiefly poorly constructed residences), which would thus kill most people, the area of lethal damage is given by $A = 100Y^{2/3}$, where Y is the yield in MT and A is in km^2 . This corresponds to an overpressure of about 4 p.s.i. ($= 2.8 \times 10^5 \text{ dyn cm}^{-2}$)³⁵.

An example of a locally devastating object would be a stony or metallic projectile 250 m diameter ($1,000 \text{ MT}$ energy), which would easily penetrate to the surface; if it struck land, it would produce a crater 5 km in diameter. (A comet of similar dimensions would fragment before reaching the surface, but its fragments would not thoroughly disperse, and the airburst would have very damaging effects on the ground). Such $1,000\text{-MT}$ events happen every $\sim 10^4 \text{ yr}$; that is, their probability of happening during a person's lifetime approaches 1% . The area of devastation is $\sim 10^4 \text{ km}^2$, or 0.002% of the Earth's area. Terrible though such an explosion would be, the damage and mortality would remain essentially local (or along coastlines, for ocean impacts), and most of the planet's population would be unaffected.

Globally catastrophic impacts. Rarer impacts of sufficiently great energy would have global consequences in addition to local devastation near the impact site. An obvious if extreme example is the K/T impact $65 \times 10^6 \text{ yr}$ ago, which disrupted the global ecosystem and is widely believed to have caused the extinction of more than half the species on Earth^{6,7}. The K/T impact of a 10-km object (nearly 10^8 MT) excavated a crater (Chicxulub) 180 km in diameter (we use the original estimates of object size and energy⁶ and of crater size, although recent evidence suggests that the event and resulting crater were larger³⁸); it apparently resulted in devastating wildfires³⁹ and changes in atmospheric and oceanic chemistry⁴⁰ as well as a dramatic short-term perturbation.

bation in climate⁴¹ produced by some 10^{16} kg of submicrometre dust injected into the stratosphere.

The K/T impact was probably the largest such terrestrial event in the last 10^8 (or even 10^9) yr. Much less massive projectiles can, however, still affect the global climate by injecting dust into the stratosphere^{41,42}. Far short of a mass extinction of species, there still could be climate changes sufficient to dramatically reduce crop yields and trigger mass human starvation⁴³. Many implications of such short-term climate change have been studied in the context of nuclear winter^{44,45}, which is analogous in some ways to a global impact catastrophe.

There is uncertainty about the threshold impact energy for global catastrophe. Quite apart from particular differences of one impact compared with another (comet/asteroid, Southern/Northern Hemisphere impact, character of 'ground zero', season, incidence angle, and so on), there are large uncertainties in the environmental consequences of impact and, even more, in the effects on human civilization. The combined uncertainty could be expressed as an uncertainty in the number of deaths that would be caused by an impact of specified energy, or it can be expressed equivalently as an uncertainty in the size of object that would kill a specified fraction of the population. For the present discussion, we choose the latter approach and *define* a globally catastrophic impact as one that would disrupt global agricultural production and lead, directly or indirectly, to the deaths of more than a quarter of the world's population (>1.5 billion people).

To appreciate the scale of the global catastrophe we have defined, we must be clear what it is not. We are dealing with the breakdown of organized agriculture but not with the collapse of most natural ecosystems. We are talking about a calamity far larger than the effects of the great world wars but far smaller than the K/T impact. We are considering a catastrophe that would destabilize modern civilization, but not an apocalypse that would threaten the survival of the human species.

Threshold for a global impact catastrophe. Having defined a global catastrophe as one that leads to the death of 25% of the world's population, we now consider what size of colliding object

(hence frequency of such catastrophes) might precipitate such consequences, and we discuss the large range of uncertainty in that size. The concept of a threshold, at the transition from local to global catastrophe, is useful. At the point where local effects of an explosion are augmented by global climatic effects, every nation and person is placed at risk, independent of location relative to the impact. Above such a threshold, the mortality (and other risks) from an impact is dramatically higher. We do not mean that the threshold is sharp; certainly the collection of different physical, environmental, and sociological responses to an impact will blur the distinction between strictly local and thoroughly global consequences (for example, tsunamis produce more-than-local but less-than-global effects). But the atmosphere's capacity to distribute sun-darkening, sub-micrometre dust from a major impact appears to be the global environmental consequence that sets in for the smallest objects⁴², and thus establishes a conceptual global threshold.

A global 'impact winter' will commence⁴² when sufficient sub-micrometre dust is injected into the stratosphere to produce optical depth >2 worldwide for a few months (until the dust falls out), and depress average land temperatures by several to perhaps 10°C or more for a period of months⁴¹ to as long as a year (due to oceanic thermal lag)⁴², causing intermittent killing frosts in mid-latitudes even in summer. This optical depth corresponds to 10^{16} g of stratospheric aerosol, or ~ 100 times as much as was lofted by any of the major volcanic eruptions of the last two centuries, including Pinatubo in 1991 and Tambora (which caused the 'year without summer' in 1816 (ref. 46)). Such an environmental shock would surely curtail agriculture production during one growing season⁴³. The repercussions would be severe, as few nations store one year's worth of food. Compounded by other effects⁷ of the impact (direct killing of millions, destruction of the ozone layer, widespread acid rain, and so on), this agricultural disaster might precipitate the collapse of global economic, social and political structures⁴⁵.

Scaling from nuclear weapons tests and the K/T event, Toon and his colleagues⁴² find that an optical depth ≈ 2 would result from a groundburst with a yield of 10^5 – 10^6 MT, corresponding to a diameter of 1–2 km for a stony object striking at 20 km s^{-1} . We adopt impact by a 1.5-km-diameter stony object with a yield of $\sim 2 \times 10^5$ MT as the nominal threshold value for a global catastrophe.

The uncertainty in the threshold energy probably exceeds an order of magnitude above or below this value. As Turco *et al.*⁴⁵ wrote, in a nuclear winter context, "the total impact . . . on biological communities is more uncertain [than are climatic effects] because little is known about the effects of physical stresses on ecosystems . . . Thus, the important issue of synergisms between various ecological stresses . . . will not be resolved, [so] the global biological impacts—hence the human impacts—of nuclear war will remain as the principal uncertainties." To reflect such uncertainties, as well as uncertainties in climate modelling and in other physical and chemical effects of an impact, we adopt (Table 1) a possible range of thresholds from 1.5×10^4 to 10^7 MT (0.6–5 km diameter). It would take a very unfavourable combination of parameters coupled with an assumption that human society is very fragile, to imagine that an object with a diameter of $\lesssim 0.5$ km could produce a global catastrophe; on the other hand, a >5 km object would⁴² create a global firestorm and so much darkness from stratospheric opacity that vision would cease—an environmental holocaust certainly exceeding any definition for the onset of global catastrophe.

Hazard analysis

The area of mortality associated with airbursts or groundbursts from objects larger than our 50-m (10 MT) threshold for penetration into the lower atmosphere depends on terrain, nature of habitation, altitude of blast, and so on. (Here we ignore iron meteoroids, which account for $<3\%$ of falls and whose potential for cratering the Earth's surface has been well studied by Shoe-

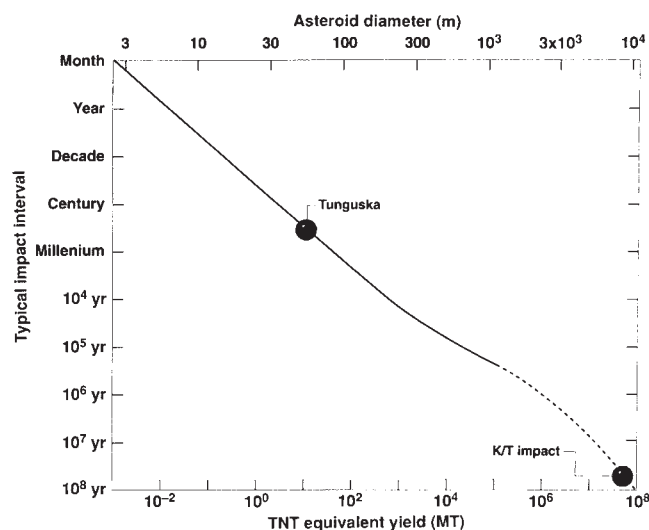


FIG. 1 Typical intervals between impacts equal to or larger than the specified yields. The solid line is Shoemaker's²¹ "best estimate," extended (dashed line) to a minimum estimate of the K/T impact energy⁶. Impact frequencies could be uncertain by a factor of 10 near 0.01 MT and by a factor of 3 for >100 MT¹⁸. The current impact rate is dominated by small-number statistics in the region of the dashed line. Equivalent asteroid diameters are shown, assuming 20 km s^{-1} impact velocity and 3 g cm^{-3} density.

TABLE 2 Fatality rates and scale of impact for three different estimates of global threshold

Type of event	Diameter of impactor	Energy (MT)	Typical interval (yr)	Deaths	World deaths per year
High atmospheric break-up	<50 m	<9	N.A.	~0	~0
Tunguska-like events	50–300 m	9–2,000	250	5×10^3	20
Large sub-global events	300–600 m	$2,000\text{--}1.5 \times 10^4$	35×10^3	3×10^5	8
	300–1.5 km	$2,000\text{--}2.5 \times 10^5$	25×10^3	5×10^5	20
	300–5 km	$2,000\text{--}10^7$	25×10^3	1.2×10^6	45
Low global threshold	>600 m	1.5×10^4	7×10^4	1.5×10^9	2×10^4
Nominal global threshold	>1.5 km	2×10^5	5×10^5	1.5×10^9	3×10^3
High global threshold	>5 km	10^7	6×10^6	1.5×10^9	250
Rare K/T scale events	>10 km	10^8	10^8	5×10^9	50

maker *et al.*⁴⁷; their results imply that they pose a minimal hazard compared with that of larger, stony objects.) We have assumed above that the zone of mortality is roughly the area of devastation (4 p.s.i. overpressure). Then, for the average world population density of ten people per km², the average number of fatalities per impact is $\sim 10^3 \times Y^{2/3}$ (sloping solid line in Fig. 2, bottom diagram, which gives expected fatalities per event for a range of asteroid sizes striking the Earth). Thus an 'average Tunguska' (occurring every few centuries) would cause $\sim 7,000$ deaths. Of course, most of the world's population is concentrated in a tiny fraction of the surface area, so most Tunguskas would strike uninhabited parts of the globe and kill few people or none at all, just as in 1908.

A projectile hundreds of metres in diameter falling in an ocean could generate tsunamis, threatening populations near the ocean rim⁴⁸. Evaluation of mortality associated with such phenomena is in its infancy; clearly it depends on the efficacy of warning and rapid evacuation. We generally neglect tsunamis below; certainly catastrophes along shorelines do not compete with catastrophes that affect the whole globe, despite concentrations of population near some coasts. Inclusion of tsunamis might raise mortality from several-hundred-metre bodies by as much as a factor of 10 (J. Pike, manuscript in preparation); such an enhancement is shown schematically in Fig. 2 over the range where tsunamis might dominate the mortality.

Above the threshold for global catastrophe, the number of fatalities is (by definition) >1.5 billion. As meteoroid size approaches that of the K/T object, virtually the entire population of the planet is likely to perish in the aftermath of the impact. For example, just above the nominal threshold of 2×10^5 MT, the expected average local casualties from the direct blast are, from the expression above, about 3 million (up to 30 million, if tsunamis were included), while indirect casualties are (by definition of the global threshold) 1.5 billion, or a factor of 500 greater. This difference reflects the different areas affected: 0.1% of the Earth's surface for the direct blast, but the entire surface for the indirect effects.

The threshold range from Table 1 is shown as a shaded area of uncertainty in Fig. 2, bottom; the boundaries of the shaded area are schematic; we do not know the shape of the transition from regional to global effects, nor the rapidity with which the expected mortality above the global threshold approaches the entire population of the planet.

Is the greatest risk from the smaller, more frequent Tunguska-like impacts that cause local or regional mortality and damage, from larger impacts near the global catastrophic threshold, or from the very rare mass extinction events? The results of integrating the fatalities multiplied by their probability of happening over segments of the impactor size–frequency distribution are shown in Table 2 for the three estimates of global thresholds

given in Table 1. A schematic representation of how these fatality rates vary with size is shown in Fig. 2, top diagram.

The smaller, frequent events larger than the 10 MT atmospheric cut-off yield annual fatality rates of about 20 deaths per year for the current world population. In reality, of course, hundreds to tens of thousands of years pass with practically no fatalities, followed by an impact that would kill thousands of people in a localized area (at average world population density) or hundreds of thousands of people if it unluckily struck a large urban area. At the high-yield extreme, the K/T impact was vastly more devastating; but even though nearly everyone would be killed by such a calamity, they are so infrequent that the annual fatality rate is only about 50 per year for the world's present population (5 billion people killed per 100 million years).

Intermediate scale objects (>300 m but smaller than the threshold for global effects) have an annual risk similar to that of smaller events, although tsunamis could augment their importance. Just above the threshold for global catastrophe, however, the fatality rate is much higher, especially if the threshold is near our nominal estimate or smaller. If 1.5 billion people are killed by the 2×10^5 MT events that occur every 5×10^5 yr or so, then the fatality rate is 3,000 per year. The annual risk per individual is about 1:1,300,000. The corresponding lifetime risk of exposure to such cosmic bombardment, for a 65-yr life span, is about 1:20,000; those are the odds that you will die as the result of the impact of a comet or asteroid near the threshold.

Perceptions of the hazard

The nature of the impact hazard is unique in human experience. Nearly all other hazards we face in life actually happen to someone we know, or at least they are reported in the news. Behaviours affecting personal health and safety, such as cigarette smoking and car driving, are overwhelmingly more important to life expectancy than are catastrophes of any sort that affect many people at once. In contrast, few people—if any at all—are known to have died in modern times from the impact of an extraterrestrial object, and the chances are very small that anyone (or everyone) will be killed by impact anytime during the next century. Thus our personal expectation of dying from an impact is extremely small, in spite of a surprisingly high level of statistical risk.

Let us consider the impact hazard compared with other hazards. Before turning to the globally catastrophic impacts, we first discuss the smaller, more frequent impacts below the global threshold. The data in Fig. 2 imply that in the next 100 yr there is about a 1% chance of an impact >1,000 MT (for example, 100 Tunguskas at once or 10% of the world's combined nuclear arsenal), which could have happened if asteroid 1989 FC had struck the Earth rather than passing by at less than twice the distance to the Moon, as it did in 1989 (ref. 49). Such an impact

could be horribly devastating, and millions of people could be killed in the unlikely event that an urban centre were struck. However, the scale of such an impact is not greatly different from that of much more common disasters. The very worst technological accidents of modern society, like the Bhopal disaster or airline crashes, affect only hundreds or, at most, thousands of people at once. The dominant cause of catastrophic death during recorded history⁵⁰ has been from natural disasters, wars, and—especially—the famines and epidemics that have killed tens of millions of people. Even in the twentieth century, there have been at least 10 separate natural disasters (during 1900–1985) that killed 100,000 to 2,000,000 people each, including 4 earthquakes, 3 floods, 2 droughts and 1 cyclone⁵¹. Tunguska-class impacts (10 MT) in populated areas are at least 100 times less frequent^{52,53} than the floods, cyclones, and earthquakes that are similarly lethal (kill 10,000 or more people). Local and regional scale impacts should not be ignored, but steps to mitigate them should be evaluated by a cost-benefit analysis that is aware of the much larger probability—indeed certainty—of future non-impact disasters of comparable or greater magnitude.

A globally catastrophic impact, however, exceeds all other disasters in that such an event could kill much of the world's population over the course of a few months or years. Not even the worst natural catastrophes compare with a global impact catastrophe. Tornadoes, cyclones, earthquakes, tsunamis, volcanic eruptions, firestorms and floods all have limits. The frequencies of the very biggest and rarest natural catastrophes are poorly known, but they must—at some magnitude—drop to zero (which is not true for impacts) because there are absolute upper limits to the amount of strain that can build up in the Earth's crust, to the power of atmospheric storms, and so on.

Furthermore, the globally catastrophic impact is qualitatively different from other more familiar hazards in its synergistic effects upon the entire planet. Even the nations most affected by the world wars of this century, in which tens of millions of people were killed and infrastructure was substantially destroyed, were able to return to productivity and a high living standard within a decade or so. This would not be expected in the aftermath of a globally catastrophic impact, because by definition it undermines all nations, which therefore cannot help each other.

One way to consider the impact hazard is to compare its statistical risk (dominated, of course, by the rare, globally catastrophic impacts, rather than more frequent, smaller impacts) with those of other hazards we face in the modern world. Table 3 gives the chances of death from selected types of accidents and natural disasters for the 'average' American obtained from the risk-assessment literature^{54–56}; they typically have uncertainties of ~30%, but some uncommon, controversial, or poorly understood hazards are uncertain by a factor of three. Risks can be much higher for selected individuals (for example, airline pilots, or backpackers in grizzly-bear country). Such risks as car accidents (but not tornadoes) are roughly similar for developed countries other than the United States; flooding and earthquakes have been far more damaging in some Asian countries than in the United States.

Each typical person (in the United States) stands a similar chance of dying in an asteroid impact as in an aeroplane crash or in a flood. By the straight odds of being killed, the impact threat is as serious as some hazards that most people take very seriously and that governments spend appreciable money to mitigate. Indeed, by the straight odds of death—the chances of a typical US citizen's epitaph reading that they died by Cause X—the impact threat is much higher than widely publicized threats from certain carcinogens, poisoning by commercial foods and pills that have been deliberately tampered with, wild animals like grizzly bears, fireworks accidents, terrorist bombs and airline hijacking.

How do we relate to such statistics? Let us examine them in a different way. Figure 3 plots some of the hazards in Table 3 as a function of the number of people killed in a single event.

For instance, most murder or accidental electrocution incidents kill one person at a time, whereas most deaths in airliner crashes occur in tragedies that can kill hundreds of people at once. In that sense, air crashes are 'rare/high consequence' events compared with 'frequent/low consequence' events in which a single individual dies at a time. Apparently, the number killed at once is viewed as being more important than total death rate, so airliner crashes receive wide attention in the news media compared with electrocutions, despite even lower risks of death. Impacts are so extreme on this scale, being exceedingly rare but extraordinarily high-consequence events, that many people feel very differently about this hazard than they do about other numerically equivalent hazards.

The overwhelmingly most likely number of people to die by a globally catastrophic impact in the foreseeable future is zero. The juxtaposition of the small probability of occurrence balanced against the enormous consequences if it does happen makes the hazard of 'impact winter' very difficult to think about. We have found⁵⁷ that people have two contrary ways of reacting to recent media discussion of the impact hazard; some dismiss it out-of-hand while others are concerned about it.

One way people think about infrequent hazards is to conclude 'it can't happen to me'. Studies of risk perception indicate that people often regard even an involuntary hazard as being of negli-

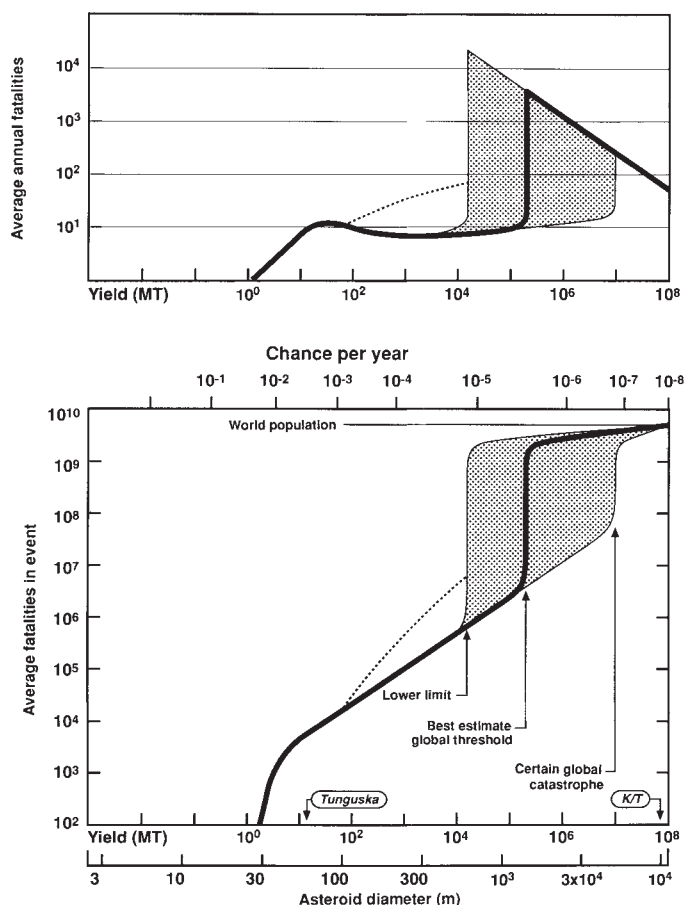


FIG. 2 Bottom, average mortality from impacts of specified yields, for average present population density on the Earth. The dashed line indicates possible enhanced mortality from tsunamis. The shaded region indicates the range of uncertainties in estimates of the threshold for global catastrophe from Table 1. Top, schematic representation of the average fatality rate (deaths per year) from impacts of different sizes. The dashed line and shaded region are as in the bottom part of the figure. Scales for associated asteroid diameters and impact probabilities from Fig. 1 are also shown.

gible concern if the risk of death per year is one chance in a million, or less⁵⁴. That is why, we suspect, some commentators have reacted to popular discussion of the impact hazard with lack of concern—even cynicism and incredulity that anyone should take the hazard seriously: the impact threat is near that psychological threshold of being dismissed because of its extreme rarity.

On the other hand, studies of risk perception by Slovic⁵⁸ suggest that many people may be inclined to regard a hazard like the impact hazard more seriously than other numerically similar hazards. (Whether people's perceptions should be regarded as a valid basis for government action, or instead dismissed as misperceptions, is a separate issue.) Slovic demonstrates that people's attitudes favouring expenditure of public funds to reduce risks are positively correlated with two factors that describe perceived attributes of hazards: the 'dreadful' nature of a hazard and the degree to which it is 'unknown.' As he defines those factors, the impact threat is both dreadful (globally catastrophic, fatal consequences, high risk to future generations, involuntary) and unknown (not yet observed, newly perceived risk). This may explain why the public and news media have recently expressed such interest in the impact hazard.

Perhaps the distinction between the two antithetical reactions to the impact hazard is that an individual worrying practically about his or her own mortality really has more important things to be concerned about, whereas an individual taking a broader, societal point of view and realizing that a large impact would be devastating, can reasonably wonder whether or not the threat can be mitigated at affordable cost.

Risk reduction and mitigation

Unlike the species extant before the K/T extinction event, human beings now can anticipate and mitigate the impact hazard. Potential impactors (comets or asteroids) can be identified and tracked, and if one is found that poses a near-term threat of impact, we have the technological capability (in principle) to intercept and deflect it. Just such a scenario was studied in 1968 as a senior class project at MIT⁵⁹ and it is the theme of a recent novel⁶⁰. NASA has sponsored several workshops to examine these options, most recently in response to a specific Congressional request (NASA Multiyear Authorization Act, 26 Sept. 1990) for options in the two areas of detection and interception.

The NASA Near-Earth Object Detection Workshop proposed a programme^{18,61}, called the Spaceguard Survey, that in 20 yr could inventory essentially all potentially threatening asteroids large enough to precipitate a global catastrophe while also providing a lead time of a few months for most incoming comets. The Survey, which would cost ~\$50 million for capital construction, and \$10 million per year for operations, could warn us of most large objects plus a fraction of smaller ones down to the atmospheric cut-off. The warning time would typically be several decades for asteroids; for long-period comets, like cousins of Swift-Tuttle, warning time might be only a few months.

The most likely result of the Spaceguard Survey would be to find no objects larger than 1 km in immediately threatening orbits, and few if any objects of any size that are predicted to strike the Earth within the next century or so. Should such an object be found, however, action could be taken to mitigate the hazard^{62,63}. Depending on the nature of the meteoroid and the lead time available, it could be deflected or destroyed⁶⁴; it would be important to avoid fragmenting it into an even more damaging swarm of pieces, which argues for stand-off nuclear explosions rather than surface or buried explosions, or kinetic energy methods. An appropriate impulse, applied by a stand-off explosion of a megaton-range nuclear device a few hundred metres above the surface, could alter the object's orbit so as to remove the immediate impact danger. Warning time of an impact by an asteroid or short-period comet is likely to be long once the Spaceguard Survey is nearly complete, but even if the warning is too short for deflection (as for long-period comets or smaller

asteroids), people near the anticipated point of impact could take shelter or be evacuated. The emphasis in a mitigation program would be on protection from the most dangerous objects, those that pose a risk of global catastrophe.

Discussion

The threat of impacts of asteroids and comets has existed since our planet's formation but it has only recently been recognized as having practical consequences for modern life. The chances that civilization might be disrupted or destroyed by such an impact are very low, but they are not zero. By some measures, they are comparable to other hazards that society takes very seriously.

Society must address the public policy issues raised by the impact hazard and proposals to deal with it. Assuming, as we do, that it is appropriate to tamper with a natural process that may have shaped evolution of life on our planet, there are questions of strategy. For example, in the case of asteroids, a comprehensive sky survey would likely provide decades or more warning, permitting decisions about specific deflection schemes to be deferred until they really are required. But a large comet might be identified on an impact trajectory with a lead time of only a few months; should we prepare in advance to deal—at great expense—with such a contingency, because of its 'unacceptable' consequences, even though it is extremely unlikely?

To date, only a small fraction of the potentially threatening objects have been discovered, but it is well within our capability to inventory most of them. If, as is extremely likely, none of the discovered objects is found to be on a collision course with Earth, then we will know that we are safe, at least from the possibility of an asteroidal impact. Even if it did little to address

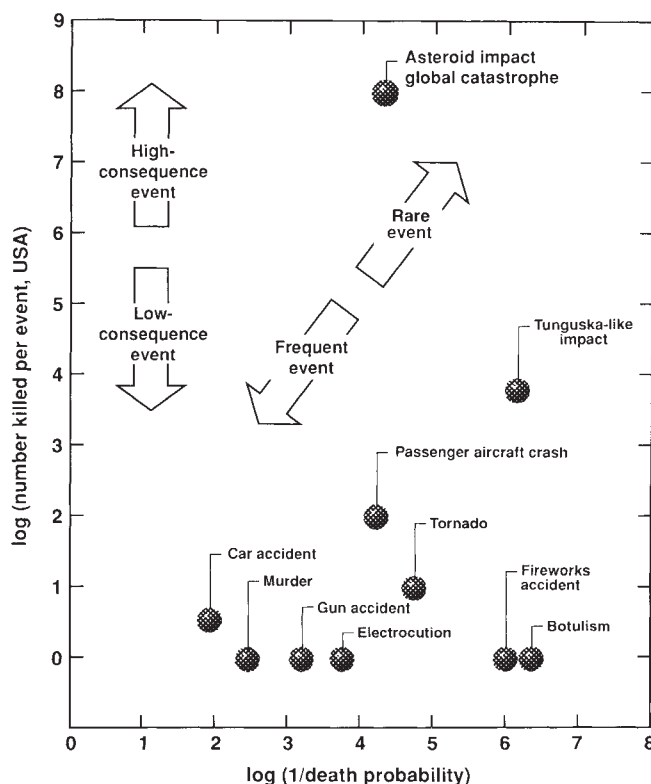


FIG. 3 Impact is an extreme case of a rare but high consequence hazard. For selected hazards (USA), the log of the number of deaths per event is plotted against log of the inverse death probability. Two impact cases are shown (global catastrophe and Tunguska-like events; the global catastrophe case is plotted assuming that the world death rate is applicable to the USA). Trends are shown for very rare and frequent events and for high-consequence and low-consequence events.

TABLE 3 Chances of dying from selected causes (USA)

Cause of death	Chances
Motor vehicle accident	1 in 100
Murder	1 in 300
Fire	1 in 800
Firearms accident	1 in 2,500
Asteroid/comet impact (lower limit)	1 in 3,000
Electrocution	1 in 5,000
ASTEROID/COMET IMPACT	1 in 20,000
Passenger aircraft crash	1 in 20,000
Flood	1 in 30,000
Tornado	1 in 60,000
Venomous bite or sting	1 in 100,000
Asteroid/comet impact (upper limit)	1 in 250,000
Fireworks accident	1 in 1 million
Food poisoning by botulism	1 in 3 million
Drinking water with EPA limit of TCE*	1 in 10 million

* EPA, Environmental Protection Agency; TCE, trichloroethylene.

the hazard posed by long-period comets, the asteroid survey alone could reduce the known hazard by at least a factor of three. If, however, we were unlucky and a body were found that would strike the Earth within the next few years or decades, it would probably be possible to divert the body so that it would miss the Earth. In such an eventuality, we expect few would argue against mounting such a mission.

But until a threatening body is actually found, we believe that preparation of a mitigation system would be premature and not cost-effective. It has been argued¹⁹ that a Spaceguard-like survey costing \$10 million per year for 30 yr may be worth the cost. Development, testing and actual implementation of a mitigation system (for example, involving launch vehicles and explosive devices) would be much more expensive. Particularly if the system involves controversial and potentially hazardous elements, such as nuclear weapons, society would surely require further expense to lessen the potential for dangerous accidents involving (or misuse of) the mitigation system itself. Therefore, building such a mitigation system would add enormous cost while it would add only modestly to our security, since it would be required only for dangers involving short lead times. For the majority of the most hazardous impactors, we would have time to develop mitigation after an object were discovered.

Although much uncertainty remains about exactly what would happen if a kilometre-scale asteroid were to strike Earth, we know more about what to expect for smaller, more frequent impacts that are likely to occur long before the 'big one'. Should we plan to do something about such events that could cause terrible regional destruction and kill millions? Their hazard is much less than the hazard of other natural catastrophes that can and do kill just as many people much more often; moreover, it would be extremely difficult to discover and inventory all of the countless potential colliding objects down to tens of metres in size. Therefore, maintenance of expensive, active surveillance and inherently risky 'space defences' against small impacts seems to be out of proportion to their threat⁶⁵.

Of course, there is perhaps a much more serious indirect threat from small impacts: misidentification of an asteroid airburst could trigger an inappropriate military response from a nuclear power in times of international tension. High-altitude airbursts with the energy of the Hiroshima bomb occur annually. It is important that such natural phenomena be understood, especially by the military decision-makers of nations with a nuclear capability. This may be the most important immediate issue raised by recognition of the impact hazard.

As public recognition of the impact hazard grows, due to more frequent discoveries and reports of 'near misses' (which is inevitable because of increased sensitivity of telescopic surveys),

such questions will increasingly fuel the hazard-mitigation debate. The impact hazard must be considered in parallel with, and balanced against, debates over society's priorities in dealing with other potential ecological disasters and hazards in general. One reason that the impact deserves careful scrutiny is that it has elements that are well-determined compared with other rare, controversial, but less understood hazards. For instance, impact rates are quite well known. Moreover, approaches to mitigation of the most likely impact threats are relatively straightforward, despite being 'high tech'. Thus, by choosing whether or not to do something about this threat from the skies, society may establish a standard against which its responses to other hazards are measured. □

Clark R. Chapman is at the Planetary Science Institute, Science Applications International Corporation, 620 N. 6th Avenue, Tucson, Arizona 85705, USA. David Morrison is at the Space Science Division, NASA Ames Research Center, Moffett Field, California 94035, USA.

1. Wetherill, G. W. in *Origins and Extinctions* (eds Osterbrock, D. E. & Raven, P. H.) Ch. 2 (Yale Univ. Press, New Haven, 1988).
2. Hartmann, W. K. et al. in *Basaltic Volcanism on the Terrestrial Planets* 1049–1127 (Basaltic Volcanism Study Project, Lunar and Planetary Inst., Pergamon, New York, 1981).
3. Grieve, R. A. F. *Meteoritics* **26**, 175–194 (1991).
4. Helin, E. F. in *Asteroids, Comets, Meteors 1991* (eds Harris, A. W. & Bowell, E.) 235–236 (Lunar & Planetary Inst., Houston, 1992).
5. Sinnott, R. W. *Sky Telesc.* **78**, (July) 30 (1989).
6. Alvarez, L., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095–1108 (1980).
7. Sharpton, V. L. & Ward, P. D. (eds) *Global Catastrophes in Earth History* (Geol. Soc. Am. Spec. Pap. No. 247, Boulder, 1990).
8. Swisher, C. C. et al. *Science* **257**, 954–958 (1992).
9. Raup, D. M. *Extinction: Bad Genes or Bad Luck?* (Norton, New York, 1991).
10. Maher, K. A. & Stevenson, D. J. *Nature* **331**, 612–614 (1988).
11. Sleep, N. H., Zahnle, K. J., Kasting, J. F. & Morowitz, H. J. *Nature* **342**, 139–142 (1989).
12. Hartmann, W. K., Phillips, R. J. & Taylor, G. J. (eds) *Origin of the Moon* (Lunar & Planetary Inst., Houston, 1986).
13. Davidson, M. *Mon. Not. R. Astr. Soc.* **92**, 46 (1932).
14. Watson, F. G. *Between the Planets* (Harvard Univ. Press, Cambridge, 1941).
15. Baldwin, R. B. *The Face of the Moon* ix (Univ. Chicago Press, 1949).
16. Krinov, E. E. in *The Moon, Meteorites, and Comets* (eds Middlehurst, B. M. & Kuiper, G. P.) Ch. 8 (Univ. Chicago Press, 1963).
17. Chapman, C. R. & Morrison, D. *Cosmic Catastrophes* Ch. 19 (Plenum, New York, 1989).
18. Morrison, D. *The Spaceguard Survey: Report of the NASA International Near-Earth Object Detection Workshop* (NASA/Jet Propulsion Lab., 1992).
19. Editorial *The Economist* **328**, 13–14 (1993).
20. Hartmann, W. K. et al. in *Basaltic Volcanism on the Terrestrial Planets* Ch. 8 (Pergamon, New York, 1981).
21. Shoemaker, E. M. *A. Rev. Earth planet. Sci.* **11**, 461–494 (1983).
22. Shoemaker, E. M., Wolfe, R. F. & Shoemaker, C. S. in *Global Catastrophes in Earth History* (eds Sharpton, V. L. & Ward, P. D.) 155–170 (Geol. Soc. Am. Spec. Pap. No. 247, Boulder, 1990).
23. Weissman, P. R. in *Global Catastrophes in Earth History* (eds Sharpton, V. L. & Ward, P. D.) 171–180 (Geol. Soc. Am. Spec. Pap. No. 247, Boulder, 1990).
24. Wetherill, G. W. *Icarus* **76**, 1–18 (1988); *Meteoritics* **24**, 15–22 (1989).
25. McFadden, L., Tholen, D. J. & Veeder, G. J. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 442–467 (Univ. Arizona Press, Tucson, 1989).
26. Everhart, E. *Astr. J.* **72**, 1002–1011 (1967).
27. Rabinowitz, D. L. *Astrophys. J.* **407**, 412–427 (1993).
28. Rabinowitz, D. L. et al. *Nature* **363**, 704–706 (1993).
29. Chyba, C. F. *Nature* **363**, 701–703 (1993).
30. Shoemaker, E. M., Bowell, E., Rabinowitz, D. L., Muinonen, K. in *Conf. Abs. Hazards Due to Comets and Asteroids* (ed. Gehrels, T.) 73 (Univ. Ariz. Press, Tucson, 1993).
31. Shoemaker, E. M. & Izett, G. A. *Lunar planet. Science* (abstr.) **23**, 1293 (1992).
32. Di Cicco, D. *Sky Telesc.* **85**, (February) 26 (1993).
33. Jacobs, C. & Spalding, R. in *Conf. Abs. Hazards Due to Comets and Asteroids* (ed. Gehrels, T.) 45 (Univ. Ariz. Press, Tucson, 1993).
34. Chyba, C. F., Thomas, P. J. & Zahnle, K. J. *Nature* **361**, 40–44 (1993).
35. Hills, J. G. & Goda, M. P. *Astr. J.* **105**, 1114–1144 (1993).
36. Zahnle, K. J. *geophys. Res.* **97**, 10243–10255 (1992).
37. Glassone, S. & Dolan, P. J. *The Effects of Nuclear Weapons* 3rd edn (US Govt Print. Off., Washington DC, 1977).
38. Sharpton, V. et al. *Science* **261**, 1564–1567 (1993).
39. Wobach, W. S., Gilmour, I., Anders, E., Orth, C. J. & Brooks, R. R. *Nature* **324**, 665–669 (1988).
40. Prinn, R. G. & Fegley, B. *Earth planet. Sci. Lett.* **83**, 1–15 (1987).
41. Covey, C., Ghan, S. J., Walton, J. J. & Weissman, P. R. in *Global Catastrophes in Earth History* (eds Sharpton, V. L. & Ward, P. D.) 263–270 (Geol. Soc. Am. Spec. Pap. No. 247, Boulder, 1990).
42. Toon, O. B., Zahnle, K., Turco, R. P., Covey, C. in *Hazards Due to Comets and Asteroids* (ed. Gehrels, T.) (Univ. Ariz. Press, Tucson, in the press).
43. Harwell, M. A. & Hutchinson, T. C. *Environmental Consequences of Nuclear War II: Ecological and Agricultural Effects* 2nd edn (Wiley, New York, 1989).
44. Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B. & Sagan, C. *Science* **222**, 1283–1297 (1983).
45. Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B. & Sagan, C. A. *Rev. Earth planet. Sci.* **19**, 383–422 (1991).
46. Rampino, M. R., Self, S. & Stothers, R. B. A. *Rev. Earth planet. Sci.* **16**, 73–99 (1988).
47. Shoemaker, E. M. et al. *Meteoritics* **25**, 409 (1990).
48. Gault, D. E., Sonett, C. P. & Wedekind, J. A. in *10th Lunar and Planetary Sci. Conf.* (abstr.) 422–424 (Lunar & Planetary Inst., Houston, 1979).

49. Chapman, C. R., Morrison, D. & Bowell, E. *Meteoritics* **24**, 258 (1989).
50. Pike, J. *Abstr. Int. Conf. on Near-Earth Asteroids* (San Juan Capistrano Res. Inst., San Juan Capistrano, 1991).
51. *World Map of Natural Hazards* (Münchener Rückversicherungs-Ges. Munich, 1988).
52. Baker, V. R., Benito, G. & Rudoy, A. N. *Science* **259**, 348–350 (1993).
53. Chinnery, M. A. & North, R. G. *Science* **190**, 1197–1198 (1975).
54. Starr, C. & Whipple, C. *Science* **208**, 1114–1119 (1980).
55. Wilson, R. & Crouch, E. A. C. *Science* **236**, 267–270 (1987).
56. Allman, W. F. *Science* **85**, (October) 38–41 (1985).
57. Morrison, D., Chapman, C., Slovic, P. in *Hazards Due to Comets and Asteroids* (ed Gehrels, T.) (Univ. Ariz. Press, Tucson, in the press).
58. Slovic P. *Science* **236**, 280–285 (1987).
59. Kleiman, L. A. (ed.) *Project Icarus* (Mass. Inst. Technol. Rep. No. 13, M.I.T. Press, Cambridge, 1968).
60. Clarke, A. C. *The Hammer of God* (Bantam, New York, 1993).
61. Morrison, D. *Mercury* **21**, 103–106 (1992); *Mercury* **21**, 110 (1992).
62. Canavan, G. H. & Solem, J. *Mercury* **21**, 107–109 (1992).
63. Canavan, G., Solem, J. C. & Rather, J. (eds) *Proc. Near-Earth-Object Interception Workshop* (Rep. No. LA-12476-C, Los Alamos National Laboratory, Los Alamos, 1993).
64. Ahrens, T. J. & Harris, A. W. *Nature* **360**, 429–433 (1992).
65. Sagan, C. *Bull. atom. Scient.* **48**, 24–26 (1992).

ACKNOWLEDGEMENTS. We thank K. Zahnle, O. Toon, A. Harris and members of the NASA NEO Detection Workshop for discussions. C.R.C. was supported by NASA Planetary Astronomy.

ARTICLES

A brain serine/threonine protein kinase activated by Cdc42 and Rac1

Edward Manser^{*}, Thomas Leung^{*}, Harfizah Salihuddin^{*}, Zhuo-shen Zhao^{*} & Louis Lim^{*†}

^{*} Glaxo-IMCB Group, Institute of Molecular & Cell Biology, National University of Singapore, Kent Ridge 0511, Singapore

[†] Institute of Neurology, 1 Wakefield Street, London WC1N 1PJ, UK

A new brain serine/threonine protein kinase may be a target for the p21^{ras}-related proteins Cdc42 and Rac1. The kinase sequence is related to that of the yeast protein STE20, implicated in pheromone-response pathways. The kinase complexes specifically with activated (GTP-bound) p21, inhibiting p21 GTPase activity and leading to kinase autophosphorylation and activation. Autophosphorylated kinase has a decreased affinity for Cdc42/Rac, freeing the p21 for further stimulatory activities or downregulation by GTPase-activating proteins. This bimolecular interaction provides a model for studying p21 regulation of mammalian phosphorylation signalling pathways.

ALTHOUGH specific functions for several Ras-related small GTP-binding proteins (p21 proteins) have been described^{1–4}, the biochemical targets of these 'molecular switches' remain largely elusive. The exception is Ras itself (most actively studied because of its link to growth control and malignancies⁵), whose GTP form can directly modulate adenylyl cyclase in yeast⁶ and mediate uncoupling of potassium channels from the G protein α -subunit, which is effected by the GTPase-activating protein p120 rasGAP in atrial cells⁷, pointing to a role for GAPs in 'downstream' signalling. In addition, genetic and biochemical studies implicate Ras in various growth-factor-stimulated signalling pathways involving both tyrosine and serine/threonine kinases; the target or mediator in these cascades appears to be the serine/threonine kinase Raf-1 (for reviews, see refs 8 and 9). We have reported a variety of mammalian GAPs for the Ras-related Rho subfamily detected by a novel overlay assay¹⁰. These p21 proteins, which apparently participate in cytoskeletal actin organization, include: Rho members (A, B and C), implicated in the assembly of focal adhesions and actin stress fibres³; two related to the yeast protein Cdc42 which participates in bud site assembly¹¹; Rac1 involved with membrane ruffling⁴ and possibly with Rac2 in neutrophil oxidase activation¹²; and growth-related p21 RhoG¹³. The GAPs for these proteins that have been identified include rhoGAP, BCR, *n*-chimaerin¹⁴, β -chimaerin¹⁵ and rasGAP-associated p190 (ref. 16). It is striking that, in addition to the conserved GAP motif, all these proteins have domains that can potentially interact with elements of established intracellular signal transduction pathways¹⁷.

It appears that the activation of p21s (by exchange of GTP) leads to signals that are both spatially and temporally limited,

because the proteins contain functionally essential lipid anchors and possess intrinsic GTPase activity (thereby terminating the signal). The Rho subfamily differs significantly with respect to rates of GTP exchange and hydrolysis. Rho, like Ras, has a slow intrinsic rate of hydrolysis¹⁸, but Rac and Cdc42 have much shorter half-lives for bound GTP. Conversely, nucleotide exchange is much slower for Rac than Ras¹⁹. *In vivo* these events are probably controlled by interacting proteins, including the 28K (*M_r* 28,000) GDP-dissociation inhibitor (GDI) which also inhibits intrinsic and GAP-stimulated GTP hydrolysis²⁰, and the *dbl* oncogene product, which stimulates nucleotide exchange for Cdc42 (ref. 21). Mutations equivalent to Gly 12 → Val in oncogenic Ras abolish the intrinsic GTPase of Rho proteins, making them unresponsive to GAPs²² and resulting in a constitutively activated phenotype for affected cells⁴. This is consistent with a requirement for bound GTP for the effector function.

We have recently characterized a novel hippocampal tyrosine kinase p120^{ACK} (ref. 23), which binds only to the GTP-bound form of Cdc42; this binding is mediated by a unique region. We have now analysed the p21 specificity and tissue localization of other proteins binding GTP-p21 proteins and purified one abundant species, a soluble 65K Cdc42- and Rac-binding protein from rat brain. This p65 protein has protein kinase activity. Its autophosphorylation and kinase activity is stimulated by binding to activated Rac/Cdc42, which thereby directly modulates the enzyme. We propose that *in vivo* activation of these Rho family members leads to the recruitment of different subsets of soluble proteins binding GTP-p21, some of which are protein kinases, that participate in pathways regulated by these p21 proteins. The p65 kinase has sequence identity to the putative protein

product of the *S. cerevisiae* *STE20* (refs 24, 25) gene, required to transmit the pheromone signal from G protein $\beta\gamma$ -subunits to downstream components, and to p120^{ACK}.

Several proteins bind to Rho-subfamily p21s

The overlay assay with [γ -³²P]GTP-p21 as substrate detects different tissue GTPase-activating proteins¹⁰ and proteins generating opposite signals. The latter correspond to proteins binding p21 (Fig. 1a). The most prominent is a group of proteins of *M_r*s 68K, 65K and 62K which bind Cdc42 and Rac1 but not RhoA, and which are abundant in brain cytosol. A 160K Cdc42-binding protein can be distinguished from the 145K protein binding RhoA and from the 150K rhoGAP (generating a lighter band). Liver contains a RhoA-binding 70K protein. These soluble binding proteins are listed in Fig. 1b; no other proteins were found in detergent extracts. The minor p55 in some tissues is derived from proteolytic cleavage of p62 (data not shown).

Binding of p65 to activated (GTP)Cdc42

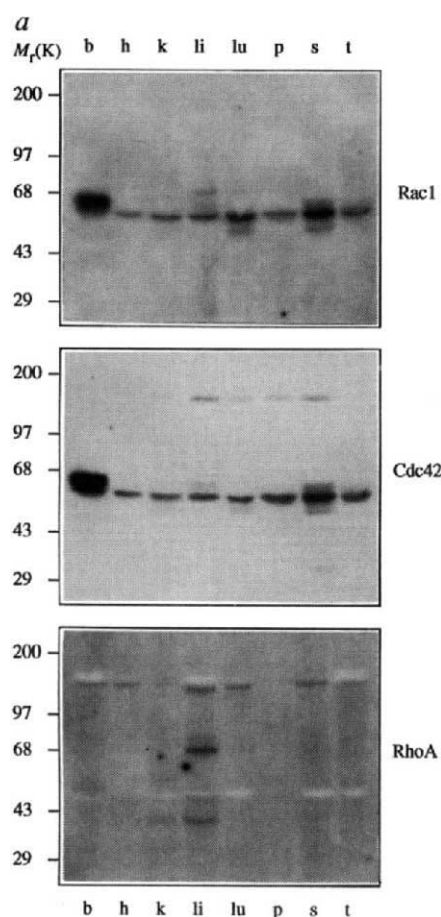
The binding proteins specifically require Cdc42 in the GTP-form and do not bind to [α -³²P]GDP-Cdc42 (data not shown). This requirement for the GTP-form is demonstrated by the interaction of partially purified brain p62–68 Cdc42/Rac-binding proteins with an immobilized fusion protein of Cdc42 with glutathione-S-transferase (GST) that has been pre-equilibrated with different guanine nucleotides (Fig. 2a). A 65K protein bound tightly to an affinity column loaded with GTP or GTP- γ S but not GDP, and weakly with the non-hydrolysable analogue GMP-PCP.

Affinity chromatography was then used to purify binding proteins. Following a three-step fractionation, the p62–68-enriched

fraction (see Fig. 4a, lane 3, for protein analysis) was loaded onto a GTP-GST/Cdc42 column, and specific proteins were eluted at pH 8.5 (Fig. 2b, lanes 1 and 2). Recovery was enhanced by repeating the affinity binding (lanes 3 and 4). Fractions 2 and 4 containing predominantly p65 were pooled. Under these conditions, p65 associates more strongly with GTP-Cdc42, allowing its selective purification. When affinity-purified p65 and flow-through proteins are separately analysed on Q-Sepharose (Fig. 2c), the latter are absent from the p65 fraction and vice versa.

p65 inhibits intrinsic and stimulated GTPases

We suggested that proteins (including p65) generated opposite signals to GAPs in the overlay assay by inhibiting the intrinsic GTPase of the p21 proteins¹⁰. Solution assays indeed demonstrate that purified p65 doubles the half-life of GTP bound to recombinant Cdc42 and Rac1 (from 8 to 15 min, and 9 to 18 min, respectively, at 20 °C; Fig. 3A, a). Because GTPase inhibition probably arises from tight association of the inhibitor with a region sensitive to the nucleotide state of the p21, this association might interfere with GAPs thought to bind near the effector loop close to the γ -phosphate of bound GTP²⁶. Indeed, p65 substantially reduces the effect of BCR, a GAP for Cdc42 and Rac1¹⁴. This effect (Fig. 3A, b) arising from inhibitor/p21 association thus requires stoichiometric amounts of p65. The p65 concentration-dependence of this GAP inhibition is biphasic (Fig. 3B, a), probably resulting from both titration of p65 against Cdc42 and competition with BCR for GTP-p21. Figure 3B, b and c, shows that having 1 μ g p65 per μ g p21 shifts the BCR-GAP response curve about 25-fold with Cdc42 and 8-fold with Rac1, which correlates with its lower affinity for the p65



<i>b</i>	<i>M_r</i> (K)	Rac1	Cdc42	RhoA	Features
	160	+/-	++	-	
	145	-	-	+	
	70	+?	-	++	Liver-specific
	68	++	+++	-	Brain-specific
	65	++	+++	-	Affinity-purified form
	62	++	+++	-	Ubiquitous
	40	-	-	+	Liver/kidney
	35	-	+	-	Spleen

FIG. 1 Detection of [γ -³²P]GTP-p21 binding proteins in soluble tissue extracts. a, Nitrocellulose filters containing cytosolic proteins separated by electrophoresis were probed with the appropriate [γ -³²P]GTP-labelled glutathione-S-transferase(GST)/p21 fusion protein. The tissues used were: b, brain; h, heart; k, kidney; li, liver; lu, lung; p, pancreas; s, spleen; and t, testis. Because of intrinsic GTPase, [γ -³²P]GTP bound to the p21 'probe' is labile, but local inhibition of GTP-p21 hydrolysis by the binding proteins enhances the overlay signal. For autoradiography, filters probed with Rac1 and RhoA and washed were exposed to Hyperfilm-MP at -20 °C overnight, or for 5 h in the case of filters probed with Cdc42. Size calibration is shown on the left. b, Summary of specificity of cytosolic p21-binding proteins identified by overlay assay with Rho subfamily members.

METHODS. Rat cytosolic proteins (250 μ g per lane) were separated on SDS-polyacrylamide (9%) gels and transferred to nitrocellulose as described¹⁰. After western blotting, nitrocellulose filters were treated with 6 M guanidinium hydrochloride in buffer Q (25 mM MES-NaOH, pH 6.5, 0.5 mM MgCl₂, 0.05 mM ZnCl₂, 0.05% Triton X-100). Filters were agitated for 5 min in this solution, which was then diluted with an equal volume of buffer Q: this was repeated five times and the filters were returned to 'renaturing' buffer¹⁰. To detect binding proteins, the filter was soaked in GAP buffer containing 1 μ g ml⁻¹ recombinant [γ -³²P]GTP-p21 and placed on 1% agarose as described¹⁰. Following incubation at room temperature for 5 min, then 10 min at 4 °C, the filters were washed three times (1 min each) in 50 mM NaCl, 25 mM MES, 5 mM MgCl₂, 0.05 Triton X-100, pH 6.5. Hyperfilm MP (Amersham) was exposed to the filters for 20 h at -20 °C. Human p21 cDNAs were obtained by PCR of human brain cDNA, subcloned into pGEX-2T and expressed as described¹⁰.

binding protein (compare with Fig. 5a). The ability of the p65 to antagonize GAPs and to inhibit GTPase points to its specific and tight binding to an important region of these molecular switches.

p65 has protein kinase activity

Concurrent isolation of the p120^{ACK} (ref. 23) led us to test for kinase activity. Purified p65 exhibited autophosphorylation *in situ* upon incubation with [γ -³²P]ATP after SDS–polyacrylamide gel electrophoresis and western blotting (Fig. 4a, lane 1). The phosphorylation pattern of proteins of the 'zinc fraction' (Fig. 4a, lane 3) suggests that other 62–68K binding proteins have this activity. Phosphorylation involved serine and threonine (see next section). As many serine/threonine kinases can be detected in this way²⁷ and only one other band (p95) is present in the 'zinc fraction', it is unlikely that the affinity-purified p65 is significantly contaminated with other kinases.

Cdc42 and Rac activate p65 kinase

Strikingly, the p65 autophosphorylation was stimulated upon co-incubation with GTP-Cdc42 and GTP-Rac during the *in vitro* kinase reaction (Fig. 4b). In addition to [γ -³²P]ATP, all reaction mixtures contained the same concentration of unlabelled GTP (as a control because GTP is a potential phosphate donor). Under these conditions, GTP-Rac was a better stimulator than GTP-Cdc42. GTP γ S-Rac1 was more effective than GTP-Rac1. Rac added to p65 kinase without prior GTP loading had very

little effect (Fig. 4b; compare lanes 8 and 12). GTP-RhoA had no effect (Fig. 4b, lane 6). In the presence of β -chimaerin, a racGAP¹⁵, GTP-Rac1 still remains effective (Fig. 4b, lane 11), demonstrating that it is not rapidly downregulated when associated with the kinase. GST/Rac fusion protein or cleaved Rac (both loaded with GTP- γ S) gave identical results (Fig. 4b, compare lanes 3 and 7). Interestingly, p21-stimulated phosphorylation resulted in the labelling of an additional 68K band (Fig. 4b, lanes 4 and 5).

After p65 autophosphorylation had been stimulated by GTP γ S-Rac (Fig. 4b, lane 3), ³²P-labelled proteins were transferred to a PVDF membrane and the 65K and 68K bands subjected to acid hydrolysis and phospho-amino-acid analysis. Figure 4c shows that both bands contain phosphothreonine and phosphoserine but no phosphotyrosine.

Autophosphorylation was then carried out for a longer period in the presence of GTP γ S-Rac1 and unlabelled ATP. The proteins were subjected to overlay assay with [γ -³²P]GTP-p21 (Fig. 5a). p21-stimulated autophosphorylation leads to conversion of p65 to a more slowly migrating species (Fig. 5a, lane 3), whereas only partial intermediate conversion is seen without p21 (Fig. 5a, lane 2). A significant decrease in Rac1 and Cdc42 binding was associated with the conversion of the p65, indicative of a lowered affinity of the phosphorylated kinase for the p21s.

p21-mediated autophosphorylation of the kinase stimulated its activity towards an exogenous substrate, myelin basic protein (MBP; Fig. 5b), which is phosphorylated on serine/threonine

FIG. 2 Selective purification of a 65K p21 binding protein by GTP-Cdc42 affinity chromatography. **a**, Binding proteins associate preferentially with the GTP form of Cdc42. Material enriched in binding protein (lane 1) prepared from brain cytosol fractionated on Sepharose-Q (1 ml protein eluted in 0.2–0.3 M NaCl) was passed over a 0.2 ml glutathione–Sephacryl column containing 8.2 mg ml⁻¹ GST/Cdc42-GTP fusion protein. Using the overlay assay with [γ -³²P]GTP-Cdc42, binding proteins are detected in the flow-through fraction (lane 2), three subsequent washes (lanes 3–5, each of one column volume), and the glutathione-eluted fraction (lane 6, two volumes) containing GST/Cdc42. Equivalent glutathione-eluted fractions from identical columns, for which Cdc42 was pre-equilibrated with different nucleotides (1 mM) are shown in lanes 7–10. **b**, Proteins of *M_r* 62–68K are specifically bound and eluted from immobilized GST/Cdc42-GTP columns. The Coomassie-blue stained gel shows proteins sequentially eluted at pH 8.5 from column A (lanes 1 and 2) and similar fractions from the flow-through fraction passed over a second column (B, lanes 3 and 4). Cdc42 (as a GST fusion protein) released from the affinity column by 5 mM glutathione is shown in lane 5. **c**, Chromatographic properties of affinity-purified p65 and flow-through Cdc42/Rac binding proteins, analysed by [γ -³²P]GTP-Cdc42 overlay. 1 ml each of purified p65, corresponding to lanes 2 and 4 in **b**, and the affinity flow-through fraction were diluted 50% and chromatographed onto 200- μ l columns of Sepharose-Q with increments of 0.1 M NaCl.

METHODS. Cytosolic proteins were subjected to a three-stage fractionation before affinity chromatography, all at 4 °C. Cytosol from 40 g rat brains was loaded onto an 80-ml Sepharose-S column, washed with 80 ml buffer Q, and the binding-protein-enriched fraction eluted with 120 ml buffer Q plus 0.25 M NaCl. This fraction in Q buffer plus 0.125 M NaCl was loaded onto a 30-ml Sepharose-Q column. After washing with 25 ml of buffer Q, proteins were eluted in 25-ml fractions with increments of 0.1 M NaCl up to 0.5 M. The pooled 0.2, 0.3 and 0.4 M fractions in 50 mM Tris-HCl, pH 7.5, were loaded onto a 10-ml chelating Sepharose (Pharmacia-LKB) column pre-saturated with ZnCl₂ according to the manufacturer's protocol, and pre-washed at pH 6. After washing with 10 ml buffer Z (100 mM NaCl, 0.5 mM MgCl₂, 0.05% Triton X-100) containing 25 mM Tris-HCl, pH 7.5, p62–68 binding proteins were eluted with 15 ml buffer Z plus 25 mM MES-NaOH, pH 6.0. This fraction in 1 mM GTP was passed over a 1-ml glutathione–Sephacryl column containing 5 mg ml⁻¹ GST–Cdc42 fusion preloaded with 1 mM GTP. After washing with one volume of buffer Z at pH 6, bound proteins were eluted in two separate volumes of buffer Z plus 50 mM Tris-HCl, pH 8.5. The procedure was repeated and the second fractions containing predominantly p65 were pooled, aliquoted and stored at –70 °C with 5% glycerol added.

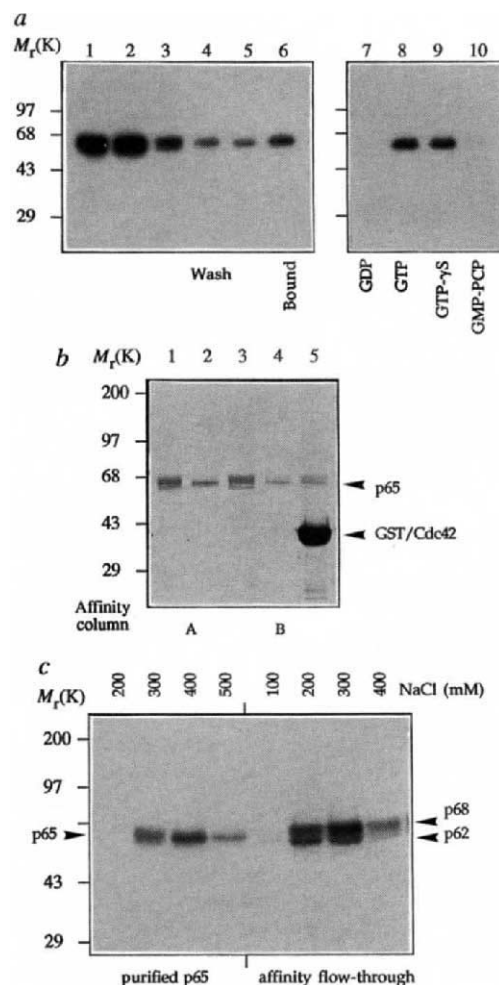
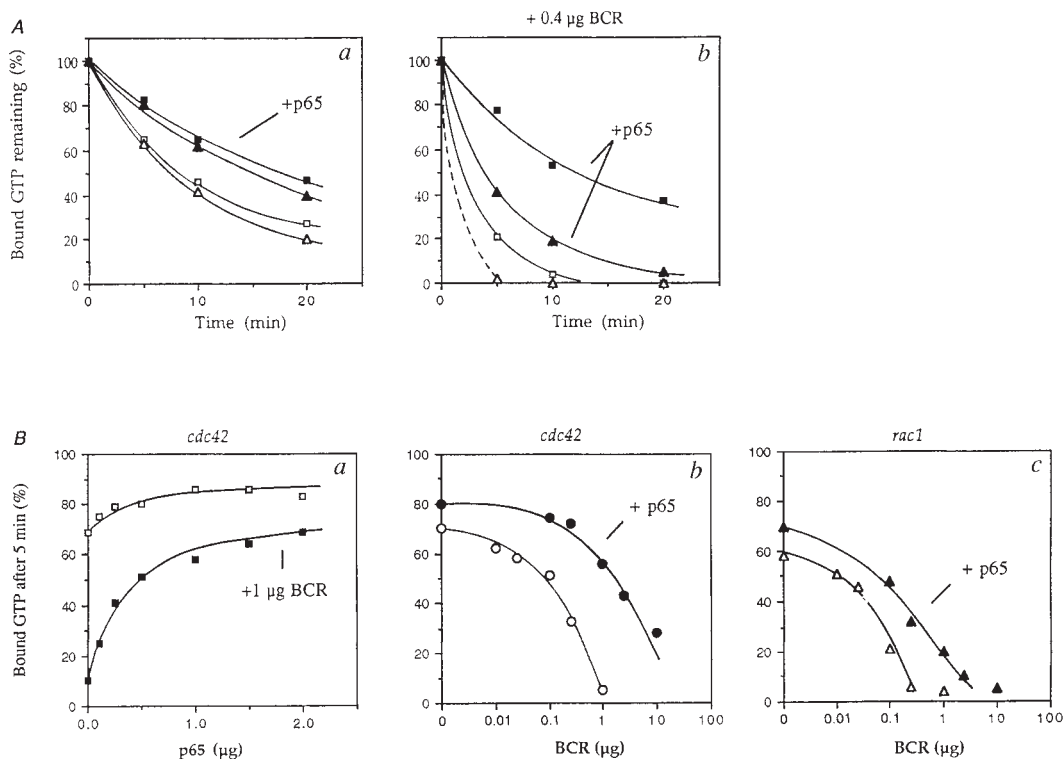


FIG. 3 The p65 binding protein inhibits intrinsic and GAP-stimulated GTPase of Cdc42 and Rac1 in a concentration-dependent manner. **A**, Hydrolysis of p21-bound GTP in the absence or presence of 0.4 μ g BCR is shown in panels **a** and **b**, respectively. 1 μ g [γ - 32 P]GTP-labelled Cdc42 ($\square$) or Rac1 ($\triangle$) was incubated at 20 °C in 100 μ l in the absence or presence of purified p65 (filled symbols). At each time-point, two 10- μ l aliquots were absorbed onto nitrocellulose and assessed for bound radioactivity; duplicates showed experimental variation of ~2%. **B**, Concentration dependence of GAP activity and its inhibition by p65 binding protein. Assays were as in **A**, with the x-axes showing the amounts of protein added per 100 μ l reaction mix (equivalent to 1 μ g p21). The effect of varying p65 concentration on GTP hydrolysis in the presence ($\blacksquare$) and absence ($\square$) of BCR is shown in **a**. The effect of varying BCR concentration on GTP hydrolysis by Cdc42 ($\circ$) or Rac1 ($\triangle$) is shown in **b** and **c** (filled symbols represent values when p65 is present). Reactions were initiated by adding GST/BCR fusion protein²³ to the radiolabelled p21 mixture (preincubated if necessary with 1 μ g purified p65 for 5 min on ice) and immediately placing at 20 °C. After 5 min, two 10- μ l aliquots were assessed for bound radioactivity. **METHODS**. p21 proteins (5 μ g) were exchanged with [γ - 32 P]GTP in 50 μ l (ref. 14); 10 μ l of this solution was added to 90 μ l 'GAP' buffer (as



described¹⁰, except that albumin was replaced with 0.05% Triton X-100 and 0.5 mM GTP was added) containing appropriate amounts of affinity-purified p65, and incubated for 5 min. GST/BCR fusion protein was added to the reaction mixture, which was immediately placed at 20 °C. Aliquots of 10 μ l were removed at appropriate times, absorbed onto nitrocellulose for 20 s and washed in cold high-magnesium wash buffer (Fig. 1 legend). Binding was quantified by liquid scintillation counting; a value taken at zero time with no added GAP corresponds to 100% bound radioactivity.

residues. Cdc42 was a better activator than Rac1 in the presence of MBP substrate (Fig. 5c). On the basis of these results we have termed the p65 a p21-(Cdc42/Rac) activated kinase (PAK). Taken together, the data suggest that the observed shift in mobility accompanies multiple phosphorylation events that are associated with activation of p65^{PAK}.

p65^{PAK} is related to yeast STE20 and p120^{ACK}

Peptides derived from gel-purified bovine brain p65^{PAK} show homology to the putative serine/threonine kinase domain of the 102K yeast protein STE20 (refs 24, 25). Using primers for the polymerase chain reaction (PCR) that were based on bovine peptide sequences, a rat brain cDNA encoding the complete sequence of p65^{PAK} was isolated (Fig. 6a, top panel). The predicted protein contains 544 amino acid residues with a calculated molecular mass of ~61K. p65^{PAK} and STE20 share ~70% identity in the serine/threonine kinase domain (Fig. 6a, middle panel). Another yeast sequence (Genbank accession no. X69322) of unknown function had 54.5% identity within the kinase domain. By contrast, no other kinase domain showed more than 35% identity^{24,25}. p65^{PAK} also has similarity with STE20 within an N-terminal sequence, part of which is related to the GTP-Cdc42 binding domain of the tyrosine kinase p120^{ACK} (bottom panel). In confirmation of the identity of p65^{PAK}, Fig. 6b shows that both [γ - 32 P]GTP-Cdc42 and Rac1 recognize the binding domain of p65^{PAK}, expressed as a GST fusion protein, giving signals comparable to that of the p120^{ACK} domain (binding only Cdc42).

Discussion

The new family of proteins interacting with activated Rac1, Cdc42 and RhoA comprises ubiquitous and tissue-specific members. Proteins corresponding to p65^{PAK} may be present at low levels in liver and spleen (Fig. 1). The abundance of these p62–68 kinases in brain correlates with its higher levels of Rac1 and Cdc42 messenger RNA²⁸, including an alternatively spliced brain-specific Cdc42 mRNA²⁹. As with GAPs, these proteins binding to Rho-subfamily p21s are detected because of their stability on SDS-polyacrylamide electrophoresis, although some require renaturation for maximal activity. A similar overlay technique detected binding of a heart cytosolic (undefined) p21 labelled with [α - 32 P]GTP to a 32K protein related to connexin³⁰. In unfractionated tissue extracts, the counteracting effects of proteins including GAPs, guanine-nucleotide-dissociation stimulators (GDS) and inhibitors (GDI) on GTP hydrolysis of test p21s preclude detection of these binding proteins by conventional solution assays. Upon fractionation and enrichment of Cdc42/Rac binding proteins, we demonstrated that they can interact with p21s in solution (which forms the basis for their purification), confirming our overlay data.

These binding proteins can be distinguished from other soluble proteins interacting with p21s, for example GDI, which also inhibits intrinsic and stimulated GTPase activity of Cdc42 (ref. 20). GDI complexes with Rho proteins that are primarily cytosolic (probably in the inactive state) in an interaction requiring p21 C-terminal isoprenylation³¹; it may play a role in cycling p21s from the membrane to the cytosol. Our experiments

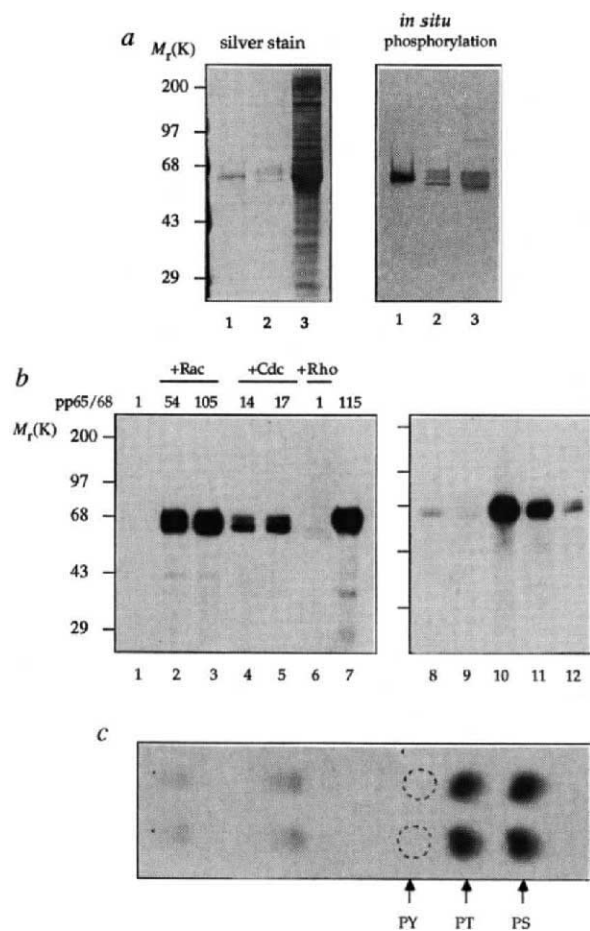


FIG. 4 Rac1 and Cdc42 can stimulate p65 autophosphorylation on serine and threonine residues. **a**, For kinase analysis *in situ*, proteins were transferred to a PVDF membrane (NEN), refolded and incubated in [γ - 32 P]ATP to allow phosphorylation. Samples of the purified p65 (200 ng, lane 1), p62–68 proteins (lane 2; early fraction from the affinity column), and the pH 6-eluted material from the chelating Sepharose column (lane 3), were run on 9% polyacrylamide gels. **b**, *In vitro* autophosphorylation of p65 (lane 1) is stimulated by GTP-Rac or GTP γ S-Rac (lanes 2 and 3), GTP-Cdc42 or GTP-Cdc42^{G12V} (GTPase-deficient⁴) (lanes 4 and 5) but not GTP-RhoA (lane 6). These were added as recombinant GST-p21 fusion proteins (2.5 μ g per 100 μ l) per 0.5 μ g kinase. Cleaved Rac1 loaded with GTP- γ S (lane 7) was as effective as GST/Rac-GTP. The level of autophosphorylation (pp65/68 relative to the control) was determined after excision and radioassay of the bands (noted above relevant lanes). In another set of experiments (unstimulated p65, lanes 8 and 9; with GTP-Rac, lane 10), equimolar racGAP (recombinant β -chimaerin¹⁵) reduced Rac stimulation of p65 autophosphorylation by ~50% (lane 11). Without pre-exchange of Rac1 with GTP (lane 12), there was little stimulation, although all final mixtures contained 0.05 mM GTP. **c**, p65 kinase is autophosphorylated *in vitro* on serine and threonine residues. Bands corresponding to 32 P-labelled p65 and p68 phosphoprotein (**b**, lane 3) were subjected to phosphoamino-acid analysis. Positions of the ninhydrin-stained phosphorylated standards are: PY, phosphotyrosine; PT, phosphothreonine; and PS, phosphoserine.

METHODS. Autophosphorylation of the kinases *in situ* was detected following electrophoresis, transfer to nitrocellulose and refolding, as for p21-binding assays (Fig. 1). The kinase reaction was carried out by incubating the filter for 30 min at room temperature in 400 μ l 'kinase buffer' (50 mM HEPES, 5 mM MgCl₂, 5 mM MnCl₂, 1 mM dithiothreitol (DTT), 0.05% Triton X-100, pH 7) containing 50 μ Ci ml⁻¹ [γ - 32 P]ATP. The filter was washed with PBS, 6 M guanidinium chloride (10 min at room temperature), then with 10% methanol, 5% acetic acid. *In vitro* kinase assays were performed in 50 μ l kinase buffer containing 10 μ Ci [γ - 32 P]ATP. For p21 activation studies, GST/p21 protein (pre-loaded with the appropriate guanine nucleotide) was added at a 5-fold excess to kinase. After incubating for 10 min at 30 °C, samples were boiled for 3 min in SDS sample buffer, run on 9% polyacrylamide gels and transferred to PVDF. Radiolabelled bands were visualized after exposure of Hyperfilm-MP to PVDF, and protein in the relevant excised bands was completely hydrolysed in 6 M HCl and analysed for radioactivity incorporated into phosphoserine, phosphothreonine and phosphotyrosine.

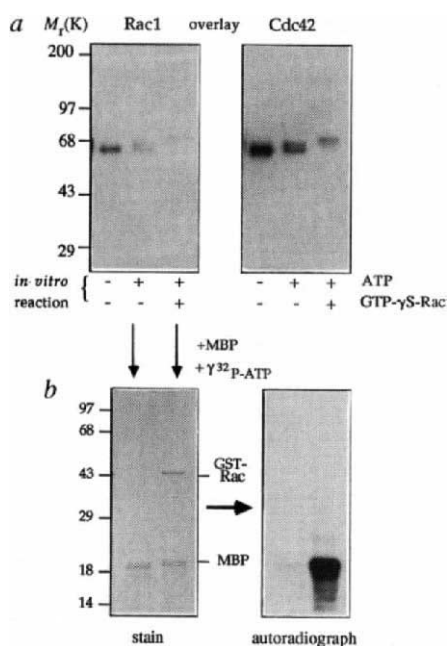


FIG. 5 The autophosphorylation of p65 kinase is associated with a decrease in p21 affinity and activation of the kinase. **a**, Purified p65 alone (1 μ g per 100 μ l; lane 1) or containing 250 μ M ATP (lane 2) with 2.5 μ g GTP γ S-Rac (lane 3) was incubated at 30 °C for 20 min. Aliquots were taken for SDS-polyacrylamide electrophoresis, western blotting and [γ - 32 P]GTP-p21 overlay as before. **b**, Samples containing non-radioactive ATP as indicated were dialysed for 2 h against kinase buffer, then supplemented with 5 μ g myelin basic protein (MBP) (Sigma) and 10 μ Ci [γ - 32 P]ATP per 50 μ l and incubated at 30 °C for 10 min. 20- μ l samples were run on a 12% polyacrylamide gel, transferred to PVDF and stained with Coomassie blue. The filter was exposed to Hyperfilm for 2 h without an intensifying screen. **c**, Activation of p65 kinase in the presence of substrate. Reaction mixtures containing 0.5 μ g p65 kinase, 5 μ g MBP and 2.5 μ M [γ - 32 P]ATP (10 μ Ci) alone (lane 1) and with either 1 μ g GTP γ S-Rac1 (lane 2) or GTP γ S-Cdc42 (lane 3), were incubated and processed as described. A 10-min exposure is shown.

involved unmodified p21s expressed in *Escherichia coli* which are unaffected by GDI and GDS^{21,32} but are acted on by GAPs and binding proteins that differ from GDI in associating only with activated p21s. This requirement is also evident in two proteins that we have cloned, the p120^{ACK} (ref. 23), and another cDNA product probably encoding the p160 Cdc42/Rac binding protein (T.L., E.M. and L.L., unpublished), which confirms it to be a general feature of this class of proteins.

We believe these binding proteins are promising candidates playing a role 'downstream' of the Rho-subfamily. This is supported by the observation that residues within the first 60 amino acids of the p21s interact with various kinases (T.L., E.M. and L.L., manuscript in preparation); this encompasses the 'effector domain' which in Ras exhibits significant conformational differences between GTP- and GDP-bound states³³. Unlike Ras, more extensive regions of Rho-subfamily proteins are required for

cellular function³⁴. The data we have presented are consistent with models proposing the existence of proteins exhibiting such GTP-p21-dependent interaction³⁵. Rabphillin 3A, which has an affinity for activated Rab3A³⁶, may represent a similar family of Rab-binding proteins. One particular problem in analysing the interaction of Cdc42 or Rac1 (with high intrinsic GTPase rates) with binding proteins *in vivo* is that multistep methods such as immunoprecipitation cannot be performed fast enough to 'trap' the GTP-p21 protein complexes. But given the specificity of the interactions and the reciprocal modulation described here for one member, the p65^{PAK}, it is likely that all the proteins we have detected (Fig. 1) are targets for the p21s.

In the light of our experimental observations, we propose a model (Fig. 6c) linking the p21 GDP/GTP cycle and the protein kinase. Conversion of the p21 to its GTP-bound form leads to its complexing with kinase. The p21, maintained as the activated

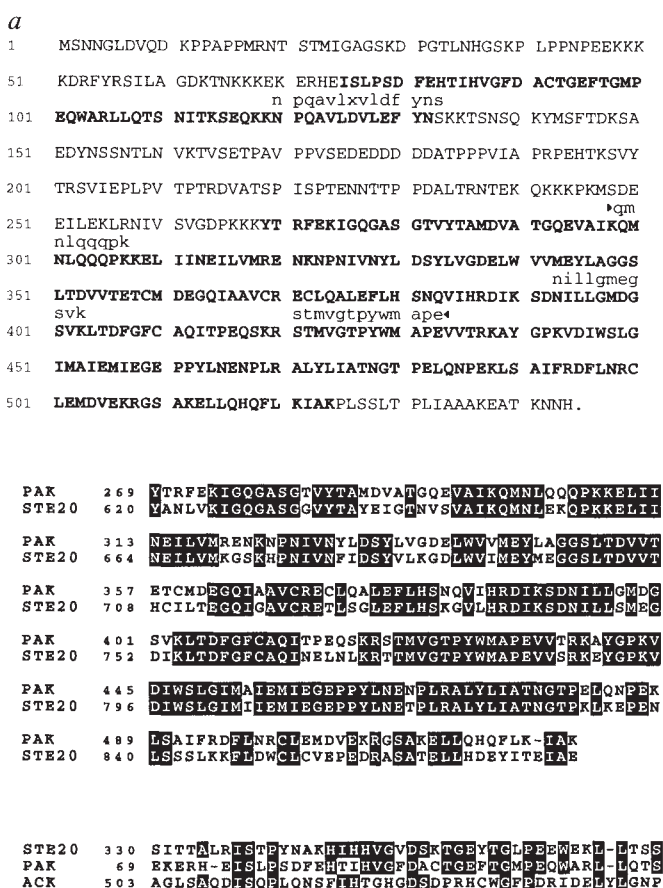
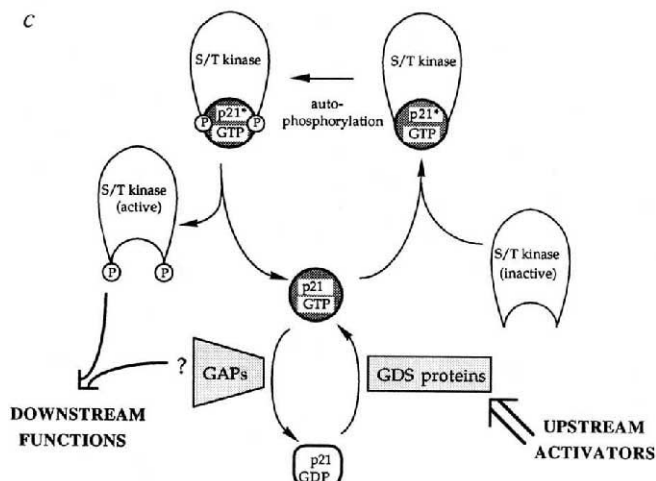
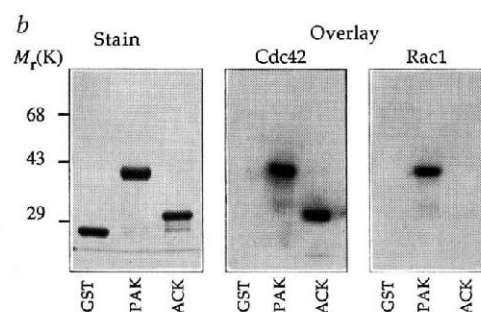


FIG. 6 Structure of p65^{PAK} and a model for its interaction with p21s. **a**, Amino-acid sequence of p65^{PAK} and comparisons with STE20 and p120^{ACK}. Top, complete sequence of p65^{PAK}; lower-case letters represent tryptic peptide sequences from purified bovine p65^{PAK}. (Peptides were coupled to Sequelon membrane through their carboxyl groups, thus generating sequence ambiguities at acidic residues, for example D/E at position 399.) Degenerate oligonucleotides corresponding to KQMNQL and WMAPEV (ends marked with an arrowhead) were used to obtain a bovine PCR product (411 base pairs) encoding a sequence identical to that enclosed by the arrowheads, except for a change at residue 376 (L to F). This DNA was used to isolate a rat brain cDNA encoding the complete sequence. Residues 269–524 (in bold) constitute the serine/threonine kinase domain whose homology to STE20 kinase domain is shown in the centre panel. Residues 75–132 (in bold) constitute a domain also present in STE20, part of which is related to the GTP-Cdc42 binding domain of p120^{ACK} (bottom panel). **b**, Cdc42 and Rac bind to the N-terminal region of p65^{PAK}. Residues 67–150 of



p65^{PAK} and 499–545 of p120^{ACK} (ref. 23) were expressed and purified as GST fusion proteins. A Coomassie-blue stained gel of these proteins is shown in the left-hand panel. Western blots on PVDF membranes containing GST (5 µg), p65^{PAK} (8 µg) and p120^{ACK} (6 µg) fusion proteins were overlaid with [γ -³²P]GTP-labelled Cdc42 (middle panel) or Rac1 (right-hand panel) as described for Fig. 1. **c**, Model for p21-mediated kinase activation. Activated p21 is thought to be generated by the action of guanine-nucleotide-dissociation stimulators (GDS), which can respond to extracellular signals. The activated Rho-family p21 binds with high affinity to p65^{PAK}. The GTP-p21 bound to kinase is in a different conformation (p21*) from the free form, as evidenced by a decrease in the intrinsic GTP hydrolysis rate. Complex formation leads to autophosphorylation of p65^{PAK} (either by intra- or intermolecular mechanisms) and stimulation of kinase activity towards its substrates. The consequently decreased affinity of phosphorylated p65^{PAK} for p21 would lead to its further preferential interaction with inactive forms of the kinase, or downregulation by GAPs.

form because of inhibition of its intrinsic GTPase, promotes autophosphorylation and activation of its kinase partner, with consequential phosphorylation of a variety of targets, which are by definition 'downstream' from the p21. Release of the p21 from the complex containing activated kinase can lead to multiple rounds of p21-mediated activation until the local pool of inactive kinase is exhausted or GAP-mediated downregulation occurs. At equimolar concentrations of GAP/p65^{PAK}, the p21 interacts preferentially with p65^{PAK} (Fig. 3). At present a role for GAPs in signalling from Rho-subfamily p21s has not been established. Activation of protein kinases by (auto)phosphorylation is well documented, for example the cAMP-dependent protein kinase³⁷ and mitogen-activated protein (MAP) kinase³⁸. With the former and S6 kinase³⁹, a mobility shift (in SDS-PAGE) similar to that of p65^{PAK} occurs upon phosphorylation-induced activation. The modulation of the interaction between p65^{PAK} and Cdc42/Rac upon autophosphorylation is similar to phosphorylation-induced down-regulation of the Cdc25 GDP-GTP exchange protein with membrane-bound Ras in yeast⁴⁰.

Since many intracellular signalling systems utilize regulatory protein kinases, it would not be surprising if the pathways of all Rho-subfamily p21s are directly linked to associated kinases. Studies of Ras-mediated signalling (primarily by cell transfection, see ref. 41) have provided much circumstantial evidence that Ras directly regulates protein kinases. In a *Xenopus* oocyte cell-free system⁴², Ras-activation of MAP kinase is mediated by a ~150K protein or protein complex that modulates the MAP-kinase kinase (MAPKK); it was suggested that the modulating factor could itself be a serine/threonine kinase. This kinase may be Raf-1, a constituent of a complex containing MAPKK and Ras⁴³, although direct association of Raf-1 with activated Ras was not demonstrated. The N-terminal regulatory region of Raf-1 does bind GTP-Ras⁴⁴⁻⁴⁶, which becomes refractory to ras-

GAP activity. Whether Ras directly affects Raf-1 kinase activity has yet to be demonstrated. The regulation of p65^{PAK} activation by Cdc42/Rac provides a model for how this might occur. In the case of RhoA, which promotes focal adhesions and stress fibres in fibroblasts³, kinases may play a role, as formation of focal adhesions involves activation of p125 focal adhesion tyrosine kinase⁴⁷. In this context, we have detected a p145 that binds activated RhoA (Fig. 1). p65^{PAK} represents the first characterized new member of a potentially extended family of p21-modulated kinases which may include p120^{ACK}. Its homology to STE20 may help in unravelling the biological role of p65^{PAK}.

In *Saccharomyces cerevisiae*, STE20 transmits the pheromone signal from G_{βγ} to downstream components including several kinases, culminating in the activation of pheromone-inducible genes (for reviews, see refs 48, 49). Whether G_{βγ} can directly activate STE20 is not known. It is possible that a p21 mediates the action of G_{βγ} in yeast; we have found that the p120^{ACK}- and p65^{PAK}-related 'regulatory' domain of STE20 (Fig. 6a, b) also binds GTP-Cdc42 (T.L., E.M., Z.Z. and L.L., manuscript in preparation). Alternatively, p65^{PAK} may represent an evolutionarily divergent, differently regulated form of the larger STE20 protein which responds separately to p21 and G_{βγ}. Given the high identity within the kinase domain (~70%), there may probably be a mammalian counterpart of at least part of the yeast signalling pathway. With respect to p21 involvement in phosphorylation pathways, it is important to establish whether the other binding proteins we have identified (Fig. 1) are kinases and the nature of their respective (target) substrates. The observed diversity of both GAPs and proteins binding the GTP form of the Rho subfamily (which appears to be more complex than for the Ras proteins) could allow a diversification of downstream paths from these p21s to allow appropriate responses of the cytoskeletal network to extracellular signals. □

Received 27 May; accepted 28 October 1993.

- Bucci, C. et al. *Cell* **70**, 715-728 (1992).
- Sluijs, P. et al. *Cell* **70**, 729-740 (1992).
- Ridley, A. J. & Hall, A. *Cell* **70**, 389-399 (1992).
- Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D. & Hall, A. *Cell* **70**, 401-410 (1992).
- Barbacid, M. *Rev. Biochem.* **56**, 779-827 (1987).
- Wang, J., Suzuki, N. & Kataoka, T. *Molec. cell. Biol.* **12**, 4937-4945 (1992).
- Martin, G. A. et al. *Science* **255**, 192-194 (1992).
- Pazin, M. J. & Williams, L. T. *Trends biochem. Sci.* **17**, 374-378 (1992).
- Ahn, G. A., Seger, R. & Krebs, E. G. *Curr. Opin. Cell Biol.* **4**, 992-999 (1992).
- Manser, E. et al. *J. biol. Chem.* **267**, 16025-16028 (1992).
- Johnson, D. I. & Pringle, J. R. *J. Cell Biol.* **111**, 143-152 (1990).
- Knaus, U. G., Heyworth, P. G., Evans, T., Curnette, J. T. & Bokoch, G. M. *Science* **254**, 1512-1515 (1991).
- Vincent, S., Jeanteur, P. & Fort, P. *Molec. cell. Biol.* **12**, 3138-3148 (1992).
- Diekmann, D. et al. *Nature* **351**, 400-402 (1991).
- Leung, T., How, B.-E., Manser, E. & Lim, L. J. *J. biol. Chem.* **268**, 3813-3816 (1993).
- Settleman, J., Albright, C. F., Foster, L. C. & Weinberg, R. A. *Nature* **359**, 153-154 (1992).
- Hall, A. *Cell* **69**, 389-391 (1992).
- Tsai, M.-H., Hall, A. & Stacey, D. W. *Molec. cell. Biol.* **9**, 5260-5264 (1989).
- Menard, L. et al. *Eur. J. Biochem.* **206**, 537-546 (1992).
- Hart, M. J. et al. *Science* **258**, 812-815 (1992).
- Hart, M. J., Eva, A., Evans, T., Aaronson, A. & Cerione, R. A. *Nature* **354**, 311-314 (1991).
- Garrett, M. D., Self, A. J., van Oers, C. & Hall, A. *J. biol. Chem.* **264**, 10-13 (1989).
- Manser, E., Leung, T., Salihuddin, H., Tan, L. & Lim, L. *Nature* **363**, 364-367 (1993).
- Leberer, E., Dignard, D., Marcus, D., Thomas, D. Y. & Whiteway, M. *EMBO J.* **11**, 4815-4824 (1992).
- Ramer, S. W. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **90**, 452-456 (1993).
- Bollag, G. & McCormick, F. A. *Rev. Cell Biol.* **7**, 601-632 (1991).
- Ferrell, J. E. & Martin, G. S. *Molec. cell. Biol.* **10**, 3020-3026 (1990).

- Didsbury, J., Weber, R. F., Bokoch, G. M., Evans, T. & Snyderman, R. *J. biol. Chem.* **264**, 16378-16382 (1989).
- Munemitsu, S. et al. *Molec. cell. Biol.* **10**, 5977-5982 (1990).
- Doucet, J.-P., Pierce, G. N., Hertzberg, E. L. & Tuana, B. S. *J. biol. Chem.* **267**, 16503-16508 (1992).
- Regazzi, R., Kikuchi, A., Takai, Y. & Wollheim, C. B. *J. biol. Chem.* **267**, 17512-17519 (1992).
- Mizuno, T. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6442-6446 (1991).
- Pai, E. F. et al. *EMBO J.* **9**, 2351-2359 (1990).
- Self, A. J., Paterson, H. F. & Hall, A. *Oncogene* **8**, 655-661 (1993).
- Hall, A. *Science* **249**, 635-640 (1990).
- Shiritaki, H. et al. *Molec. cell. Biol.* **13**, 2061-2068 (1993).
- Steinberg, R. A., Cauthron, R. D., Symcox, M. M. & Shuntoh, H. *Molec. cell. Biol.* **13**, 2332-2341 (1993).
- Payne, M. D. et al. *EMBO J.* **10**, 885-892 (1991).
- Grove, J. R. et al. *Molec. cell. Biol.* **11**, 5541-5550 (1991).
- Gross, E., Goldberg, D. & Levitzki, A. *Nature* **360**, 762-765 (1992).
- Roberts, T. M. *Nature* **360**, 534-535 (1992).
- Itoh, T. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 975-979 (1993).
- Moodie, S. A., Willumsen, B. M., Weber, M. J. & Wolfman, A. *Science* **260**, 1658-1661 (1993).
- Zhang, X.-f. et al. *Nature* **364**, 308-313 (1993).
- Warne, P. H., Viciano, P. R. & Downward, J. *Nature* **364**, 352-355 (1993).
- Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. *Cell* **74**, 205-214 (1993).
- Schaller, M. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5192-5196 (1992).
- Marsh, L., Neiman, A. M. & Herskovitz, I. *Rev. Cell. Biol.* **7**, 699-728 (1991).
- Errede, B. & Levin, D. E. *Curr. Opin. Cell. Biol.* **5**, 254-260 (1993).

ACKNOWLEDGEMENTS. We thank the Glaxo Singapore Research Fund for support, C. Pallen and J. Cairns for help and advice with protein kinase assays, and R. Philips for peptide analysis.

Exciton strings in an organic charge-transfer crystal

M. Kuwata-Gonokami*, N. Peyghambarian†, K. Meissner†, B. Fluegel†, Y. Sato*, K. Ema*, R. Shimano*, S. Mazumdar†‡, F. Guo†, T. Tokihiro*, H. Ezaki* & E. Hanamura*

* Department of Applied Physics, University of Tokyo, Tokyo 113, Japan

† Optical Sciences Center, ‡ Department of Physics, University of Arizona, Tucson, Arizona 85721, USA

COLLECTIVE excitations resulting from many-body Coulomb interactions have been studied extensively in the solid state¹: for example, the exchange interaction between the electrons in two excitons (bound electron-hole pairs) can bind the excitons together, forming a biexciton. At the other extreme, if the number of excitons is sufficiently large ($\sim 10^6$), they can condense into a degenerate 'liquid' phase known as an electron-hole drop. But in conventional semiconductors, intermediate bound states, consisting of more than two excitons, are not formed. We show here, both theoretically and experimentally, that bound states of multiple excitons can form in the organic charge-transfer solid anthracene-(pyromellitic acid dianhydride). Coulomb interactions along the one-dimensional stacks of this material can stabilize trains of several charge-transfer excitons, and we refer to the resulting collective excitations as exciton strings.

The charge-transfer (CT) crystal, anthracene-PMDA (pyromellitic acid dianhydride), is a prototype quasi-one-dimensional neutral mixed-stack solid^{2,3}. The donor (D) anthracene molecules and the acceptor (A) PMDA molecules (Fig. 1) are arranged in one-dimensional stacks with alternating D and A. CT excitation promotes an electron from the highest occupied molecular orbital (HOMO) of D to the lowest unoccupied molecular orbital (LUMO) of A, creating a D^+A^- unit. These systems are described within the extended Hubbard hamiltonian^{4,5}, H , where

$$H = \sum_{l,\sigma} (-1)^l \varepsilon c_{l\sigma}^\dagger c_{l\sigma} + t \sum_{l,\sigma} (c_{l\sigma}^\dagger c_{l+1,\sigma} + c_{l+1,\sigma}^\dagger c_{l\sigma}) + \frac{1}{2} \sum_{l,m} V_{lm} \rho_l \rho_m \quad (1)$$

where we adopt the convention that the D(A) molecules occupy odd(even) sites. Here $c_{l\sigma}^\dagger$ creates an electron of spin σ at site l , $2\varepsilon = I_D - A_A$, where I_D is the ionization energy of D and A_A is the electron affinity of A, t is the hopping matrix element and $V_{lm} = V_{|l-m|}$ is the Coulomb interaction between charged sites, $|l-m|$ lattice constants apart. The charge densities $\rho_l = z - \sum_{\sigma} c_{l\sigma}^\dagger c_{l\sigma}$, where $z = 2(0)$ for a D(A) site. In equation

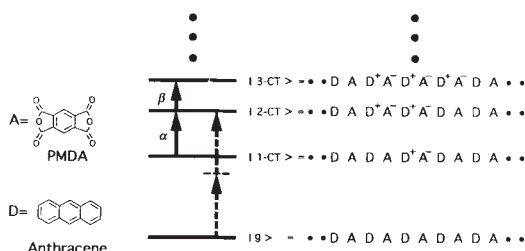


FIG. 1 Diagram showing the ground state (g) and the $n=1, 2$ and 3 exciton states (shown as 1-, 2- and 3-CT) in a neutral CT solid. Also shown are the associated photoinduced absorptions α and β to the 2-exciton and the 3-exciton states respectively, and direct TPA (dashed arrow) to the 2-exciton state. The molecular structures of the donor anthracene molecule and the acceptor PMDA molecule are also shown.

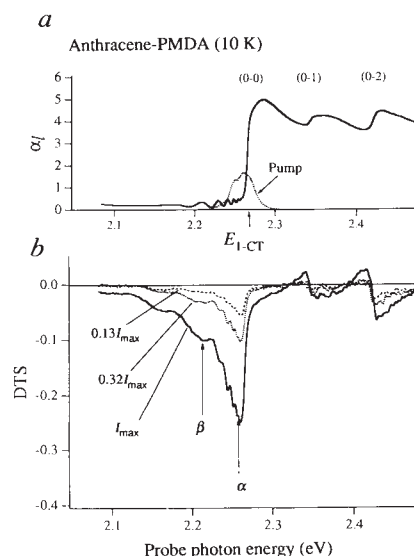


FIG. 2 a, Linear absorption (α) of anthracene-PMDA with the electric field vector parallel to the stack axis. The pump frequency coincides with the 0-0 linear absorption. The laser system is based on a colliding pulse mode-locked laser amplified by a diode-pumped Nd:YLF laser running at 1 KHz. The pump pulses were tunable in the range, 2.21–2.29 eV corresponding to wavelengths of 560 and 540 nm, respectively. b, The differential transmission spectrum (DTS) at three different pump intensities; $I_{\max}$ corresponds to the maximum pump fluence of 1 mJ cm^{-2} . The fine structure due to Fabry-Perot interference effects seen in a are easily distinguished from the intrinsic signals by means of their dependence on sample thickness.

(1), the first term corresponds to the energy difference between the HOMO of D and the LUMO of A, the second term describes charge-transfer from D to A and the last term is the Madelung electrostatic energy.

Although our quantitative calculations are for nonzero t , a simple physical picture for formation of multiexciton states can be obtained by considering the atomic limit $t \rightarrow 0$, where electron hopping is ignored, and individual energy states are described almost entirely by single configurations. The ground state is the perfectly neutral configuration ... DADADA ... The excitation energy E_{1-CT} to the 1-exciton state ... DAD^+A^-DA ... is given by

$$E_{1-CT} = 2\varepsilon - V_1 \quad (2)$$

where V_1 is the nearest-neighbour Coulomb interaction. The n -exciton state consists of n consecutive D^+A^- units, as shown in Fig. 1, and can be reached by optically exciting the $(n-1)$ -exciton. The energy of the n -exciton is,

$$E_{n-CT} = 2n\varepsilon - V_1 \cdot M(n) \quad (3)$$

with

$$M(n) = \sum_{k=1}^{2n-1} (-1)^{k+1} \frac{(2n-k)}{k} \quad (4)$$

for a $1/r$ Coulomb potential. From equations (3) and (4), the energy difference between the n -exciton and the $(n-1)$ -exciton decreases with n . The n -exciton states should therefore be detectable as photo-induced absorptions (PA) on the low-energy side of the exciton resonance in pump-probe experiments. A schematic diagram of the optical transitions between the n -exciton states is shown in Fig. 1.

High-purity thin ($< 50 \mu\text{m}$) single crystals of anthracene-PMDA were grown by sublimation techniques. Figure 2a shows the linear absorption spectrum of our sample at 10 K. The lowest-energy peak corresponds to the 0-phonon 1-exciton,

whereas the higher-energy peaks result from vibronic couplings with PMDA molecular vibrations⁶. The sample quality was evaluated by measuring the absorption spectrum at 2 K with a spectral resolution of 0.01 nm. The width of the 0-0 line, centred at 2.2713 eV, is <0.3 meV.

A pump-probe technique was employed to detect transitions between the multi-exciton states. The pump pulse generates the excited state, while the time-delayed probe pulse monitors transitions between the excited states. The differential transmission spectrum (DTS) measures the difference between the probe transmission in the absence and presence of the pump. The pump photon energy was centred at the 0-phonon line (2.2713 eV), in order to resonantly excite the $n=1$ exciton state, as shown in Fig. 2a. The DTS with a 1 ps probe delay at 10 K is shown in Fig. 2b. At low pump intensity, strong PA (denoted by α) is observed just below the exciton 0-phonon energy, whereas at higher intensity, an additional PA, denoted by β , is observed at even lower energy. The time-dependence of the PA signals, shown in Fig. 3, indicates that the recovery of the α -band is very slow and is almost unchanged up to 150 ps, whereas the β -band has a fast component with a decay time of 50 ps.

Exact numerical calculations for nonzero t , based on equation (1), were performed for a periodic ring consisting of 7 DA units. The existence of a mirror-plane of symmetry, the longest diagonal of the ring, dictates that all eigenstates are of even or odd parity. The ground state is even, and 1-photon transitions are allowed only to odd-parity states, whereas 2-photon transitions occur to even-parity states. The calculated energy spectrum, consisting of both 1- and 2-photon states, is shown in the inset of Fig. 4. As well as the exciton states shown in Fig. 1, realistic values of t also give continuum states, which consist of separated excitons as well as completely separated D^+ and A^- ions. The number of exciton levels within each parity subspace are 1, 2 and 5 for $n=1, 2$ and 3, respectively. The larger number of exciton states for higher n is due to the different spin couplings between the multiple D^+ and A^- radical ions. The most significant result of the calculation is the stability of the 2-exciton as well as the 3-exciton, as measured by the energy difference between the highest exciton and the lowest continuum state for each n . These energy separations are the binding energies of the n -exciton strings, and are indicated by the arrows in the inset of Fig. 4. For $V_1=0.5$ eV, a value consistent with the charge-transfer gap energy and the d.c. Stark effect measurement³, the 3-exciton is actually slightly more stable than the 2-exciton. Numerically obtained exact dipole couplings between all exciton and continuum states were used to calculate the DTS within fifth-order perturbation theory (M. K.-G. *et al.*, manuscript in preparation). The intensity dependence of the calculated DTS, shown in Fig. 4, is nearly identical to the experimental DTS; the stronger pump intensity gives both the α -band corresponding to the 2-exciton and the β -band corresponding to the 3-exciton. The structure between the α -band and the β -band, seen for strong intensities both in the experimental and in the calculated DTS, results from transitions from the $n=2$ excitons to the $n=$

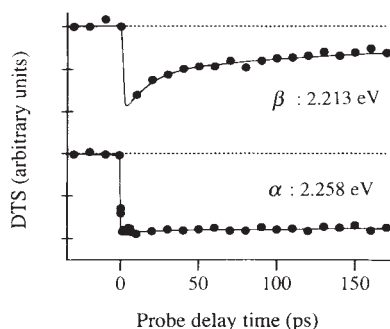


FIG. 3 The time dependence of the DTS signals α and β .

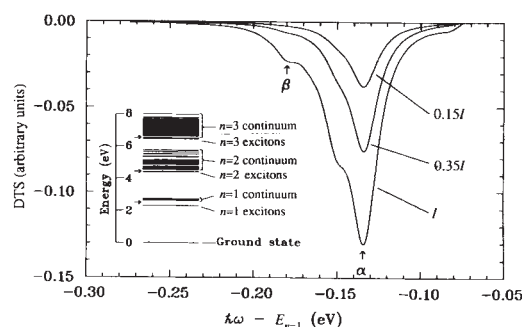


FIG. 4 Calculated DTS from equation (1) for different pump intensities I . The inset shows the energy spectrum, with the parenthesis indicating the continua for various n , and the arrows indicating the exciton binding energies. The calculations were done for a periodic ring of 14 sites, with $t=-0.1$ eV, $V_1=0.5$ eV, $V_{|m|}=V_1/m$ and $\varepsilon=1.38$ eV. DTS calculations for other reasonable parameters $t=-0.15$ and -0.2 eV and various V_1 and ε , gave similar results. The above magnitudes of t and V_1 are typical for CT solids, whereas ε was chosen to fit the experimental energy of the 1-exciton. The larger energy shifts in the calculated DTS compared to the experiment are finite-size effects.

3 continuum, and is due to the long-range nature of the Coulomb interaction.

Accordingly, we assign the α -band in Fig. 2b to the transition from the 1-exciton to the 2-exciton state, and the β -band to the transition from the 2-exciton to the 3-exciton. The α -absorption, expected to last for the 1-exciton lifetime, shows no decay signal for time delays up to 150 ps. This is consistent with the nanosecond lifetime of the 1-exciton, as measured by fluorescence (M. K.-G. *et al.*, manuscript in preparation). Independent confirmation of this assignment of the α -band is obtained by direct 2-photon absorption (TPA) to the 2-exciton (Fig. 1). The 2-exciton state of even parity may be reached directly from the ground state by simultaneous absorption of two photons; the same 2-exciton state seen in DTS is obtained by absorption from the odd-parity 1-exciton. The pump was detuned to 2.233 eV, because one-half the energy of the 2-exciton is smaller than the energy of the 1-exciton. This large detuning from the 1-exciton absorption also ensures a low density of 1-excitons. A strong transient PA signal that temporally follows the pump beam is observed. During the temporal overlap with the probe pulse, one photon from each pulse is absorbed to create the 2-exciton by direct TPA.

The second induced absorption band, β , in Fig. 2b is attributed to the transition between the 2-exciton and the 3-exciton. The β -band is prominent only at strong pump intensities, as is expected for processes higher than third order. The faster recovery rate of the β -signal in Fig. 3 is also expected, as the 2-exciton can decay into the $n=1$ states.

The exciton strings demonstrated here are different from the Bose-condensed exciton state and the electron-hole liquid drop. Bose condensation is a macroscopic occupation of a single momentum state, and the electron-hole drop in Ge and Si is a liquid phase of a large number (10^6) of excitons. CT solids are known to exhibit pressure⁻², temperature⁻² and photo-induced⁷ neutral-to-ionic phase transitions. Spectroscopy of multi-exciton states may be able to shed light on the photoinduced phase transition, which is not well understood at present. □

Received 21 July; accepted 16 November 1993.

1. Peyghambarian, N., Koch, S. W. & Mysyrowicz, A. *Introduction to Semiconductor Optics* (Prentice Hall, Englewood Cliffs, 1993).
2. Torrance, J. B. *et al.* *Phys. Rev. Lett.* **47**, 1747-1750 (1981).
3. Haarer, D., Philpot, M. R. & Morawitz, H. *J. chem. Phys.* **63**, 5238-5245 (1975).
4. Garito, A. F. & Heeger, A. J. *Acc. Chem. Res.* **7**, 232-240 (1974).
5. Soos, Z. G., Kuwajima, S. & Harding, R. H. *J. chem. Phys.* **85**, 601-610 (1986).
6. Brillante, A. & Philpot, M. R. *J. chem. Phys.* **72**, 4019-4030 (1980).
7. Koshihara, S., Tokura, Y., Mitani, T., Saito, G. & Koda, T. *Phys. Rev. B* **42**, 6853-6856 (1990).

Efficient phase-matched second-harmonic generation of blue light in an organic waveguide

Thomas L. Penner, Hubert R. Motschmann, Nancy J. Armstrong, Matthew C. Ezenyilimba & David J. Williams

Imaging Research and Advanced Development, Eastman Kodak Company, Rochester, New York 14650-2021, USA

THE development of materials for the efficient frequency-doubling of low-power semiconductor lasers is a technologically important goal for which several important challenges remain. The ease of processing and high intrinsic optical nonlinearity of organic materials^{1,2} makes them attractive for such applications, and it has been demonstrated that molecular-assembly techniques can achieve the macroscopic non-centrosymmetric order required for bulk nonlinear optical activity^{3,4}. But to achieve efficient power conversion, a promising material must also exhibit low attenuation of light at high power densities, and match phase velocities over relatively long distances for light at the fundamental and second-harmonic frequencies⁵. Here we report the fabrication by Langmuir–Blodgett deposition of a low-loss optical waveguide in which precise control of the film thickness, together with inversion of the nonlinear susceptibility across the film, are used to simultaneously achieve phase matching and improve the optical-field overlap between the propagating (fundamental and second-harmonic) waveguide modes. The resulting structure converts low-power near-infrared laser light efficiently to blue light.

The materials and fabricated film structure used are shown in Fig. 1. A control film of similar thickness but without structural inversion was also fabricated. The Langmuir–Blodgett (LB) films were prepared from preformed polymers. Synthesis of the chromophore-containing polymer and details of the LB film fabrication are reported elsewhere⁶. The short absorption wave-

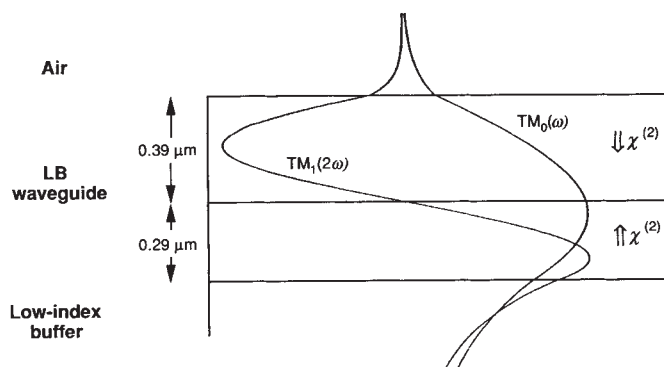


FIG. 2 Optical layer structure of the waveguide prepared by LB deposition. Thickness dimensions were determined by numerical computer modelling of the conditions needed for TM mode-conversion phase-matching at 900 nm, based on experimental refractive indices of the control film measured by prism coupling at different laser wavelengths (441.6, 454.5, 457.9, 465.8, 472.7, 488.0, 501.7, 514.5, 632.8, 830 and 1,064 nm). Electric-field distributions for both the zero-order mode at a fundamental wavelength of 900 nm and for the first-order mode at a second-harmonic wavelength of 450 nm are shown. The calculated location of the node in the first-order field was used to determine the point at which to reverse the dipping direction in the fabrication of the inverted-structure film, assuming that no change in refractive index results from this process. The calculated power confinement in the waveguide was >80% for the depicted modes.

length of the nonlinear chromophore ($\lambda_{\text{max}} = 315 \text{ nm}$) ensures low waveguide absorption losses. Macroscopic non-centrosymmetry was obtained by alternately depositing the active material with a low-volume inert spacer polymer, poly(*t*-butyl methacrylate; Aldrich)⁷. Although the side-chain chromophores have substantial order with an average tilt angle of 42° (calculated from second-harmonic generation (SHG) polarization data⁸) the repeat-unit area on the water surface is twice that for tightly packed fluorocarbon chains. This partial disorder in the perfluorocarbon 'tails' suppresses formation of domains of optical dimensions in the spread monolayer, as confirmed by Brewster-angle microscopy⁹, resulting in low optical scattering of the waveguide, in contrast to other LB films. Optical attenuation, measured by monitoring the light scattered out of the waveguide with a calibrated charge-coupled-device array, was between 1.5 and 2.0 dB cm^{-1} at 457.9 nm for both TE (transverse electric) and TM (transverse magnetic) polarizations, superior to any previously reported LB films and sufficient for practical applications. Prism coupling angles were used to calculate the LB waveguide film indices as a function of polarization and wavelength, although even the high degree of accuracy attained (± 0.0005) leaves some uncertainty in calculating the exact phase-matching conditions¹⁰.

Preliminary nonlinear characterization of the LB films was carried out using a simple transmission geometry with a pulsed Nd/YAG-pumped dye laser (20 mJ per pulse, 10 ns pulse width, 860 nm) incident on the film at 45° . From SHG intensities the nonlinear susceptibility $\chi^{(2)}$ of the film with structural inversion was 5.8 times less than the control, close to a predicted reduction factor of 6.6 calculated assuming additive susceptibilities of the individual layers. This confirms that the sign of the susceptibility is actually reversed in the former case and also that the layer structure is quite regular. Lack of detectable *s*-polarized SHG (that is, polarized perpendicular to the plane defined by the transmitted beam and the normal to the film plane) indicates $C_{\infty v}$ symmetry about an axis normal to the film plane⁸. Both the polarization ratio and the absolute SHG intensities of the control film have remained unchanged ($\pm 10\%$) over four

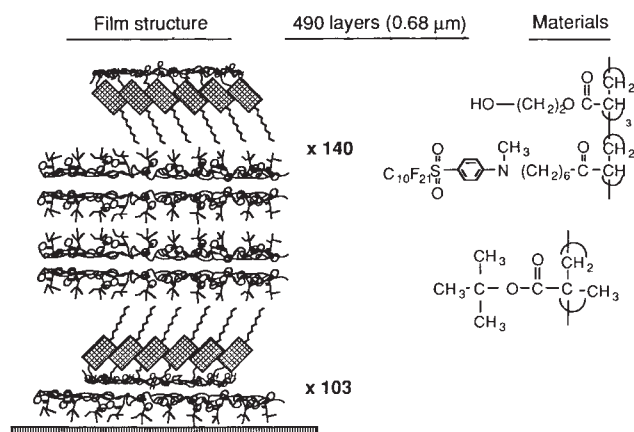


FIG. 1 Schematic representation of the molecular structure of the alternate-layer LB film with directional inversion, indicating the number of deposited layers. Chemical structures of the materials used are shown. Because the refractive index of the LB film is ~ 1.47 , a $2.0\text{-}\mu\text{m}$ -thick spin-coated layer of a fluorohydrocarbon polymer of low refractive index (1.40) was interposed to provide isolation from the Pyrex substrate and confinement of zero- and first-order modes of the fundamental and second-harmonic wavelengths respectively. The LB film was $0.68 \mu\text{m}$ thick, measured by stylus profilometry.

months. The nonlinear susceptibility of this film at 860 nm was estimated to be 8 and 10 pm V⁻¹ for $\chi^{(2)}_{xxz}$ and $\chi^{(2)}_{zzz}$ respectively, calculated from the oriented gas approximation⁷.

Mode conversion as described in the literature was used to match fundamental and second-harmonic phase velocities¹¹. By careful adjustment of the waveguide thickness the dispersion of the medium allows different modes at the two wavelengths to have the same effective refractive index n_{eff} . Using the concept proposed several years ago of reversing the sign of $\chi^{(2)}$ at the thickness where the electric-field distribution of the first-order mode undergoes a sign reversal should allow the zero-order mode of the fundamental and the first-order of the second harmonic to overlap constructively across the waveguide as shown in Fig. 2 (refs 12, 13). Simulations show that the efficiency of this mode-conversion phase-matching scheme is highly sensitive to the control of the thickness parameters, especially the overall waveguide thickness¹⁴. We estimate that for a phase-matched interaction length of 1.5 mm, a thickness error of 6 nm or ~1% results in an efficiency loss of 50%. For organic films, only a molecular fabrication technique can sustain such accuracy. Mode-conversion phase-matching is ~2.5 times more efficient than quasi-phase-matching (QPM), where the phase mismatch is periodically corrected by changing the sign of $\chi^{(2)}$ along the waveguide every coherence length (~3 μm) (ref. 15). The fundamental input wavelengths at which phase-matching is expected are calculated from this model and plotted in Fig. 3 as a function of the waveguide thickness and light polarization. As a result of the birefringence of the LB film, the phase-matching condition is satisfied at two fundamental wavelengths of different polarizations for a given film thickness, resulting in the two curves in Fig. 3. (For $C_{\infty v}$ symmetry TM- but not TE-polarized SHG is allowed.) Therefore, unlike QPM, this scheme offers the possibil-

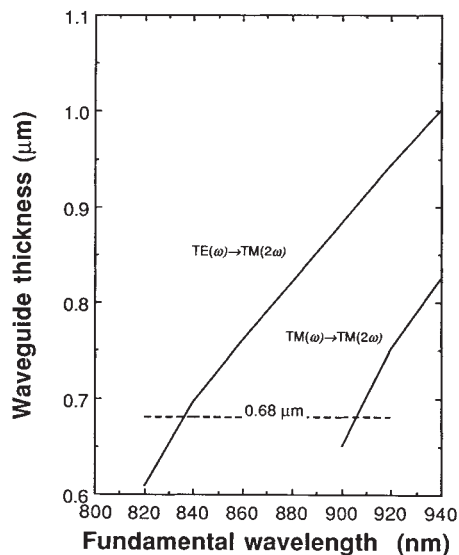


FIG. 3 Theoretical dependence of the fundamental wavelength for phase-matching on film thickness for the structurally inverted film, with the inversion fixed at 42% of overall thickness. Although four phase-matching wavelengths are theoretically possible for the different polarization combinations, only those with TM-polarized SHG are symmetry-allowed. At the film thickness fabricated, the predicted phase-matching wavelengths are: $\text{TE}_0(\omega) \rightarrow \text{TM}_1(2\omega) = 837 \text{ nm}$; $\text{TM}_0(\omega) \rightarrow \text{TM}_1(2\omega) = 906 \text{ nm}$. These calculations are very sensitive to refractive-index variations, so that a shift in refractive-index dispersion between ω and 2ω of the order of 0.0002 changes the wavelength by ~10 nm. There is thus substantial uncertainty in these simulations because our refractive-index measurements are only accurate to ± 0.0005 . The termination of the curves at low thickness is determined by the capability of the waveguide to support a first-order mode at the second-harmonic wavelength.

ity of phase-matching more than one wavelength with only changes in the input polarization.

To demonstrate phase-matching, the second-harmonic (2ω) output from the end of the 1-cm inverted-structure waveguide was monitored as a function of the wavelength of the prism-coupled fundamental (ω) input light. The SHG peaks in Fig. 4 at 819 and 888 nm confirm the phase-matching for both $\text{TE}_0(\omega) \rightarrow \text{TM}_1(2\omega)$ and $\text{TM}_0(\omega) \rightarrow \text{TM}_1(2\omega)$ polarizations. Although sought, TE-polarized SHG was not observed, consistent with the system's symmetry. The SHG intensity varies quadratically with the fundamental power. Agreement with the predicted wavelengths is good, given the limitations on refractive index accuracy. The widths of the peaks decrease with increasing refractive index dispersion of the waveguide or the phase-matched interaction length L . The form of the function describing these dependencies is³: $\text{SHG} \propto \sin^2(L\Delta k/2)/(\Delta k/2)^2$, where $\Delta k = 4\pi[n_{\text{eff}}(2\omega) - n_{\text{eff}}(\omega)]/\lambda_\omega$. We fitted the peaks in Fig. 4 to the theoretical function, with L as the only adjustable parameter¹⁶. Values of $L = 0.16$ and 0.12 mm were obtained at 819 and 888 nm, respectively. SHG intensity at 410 nm of ~1 μW

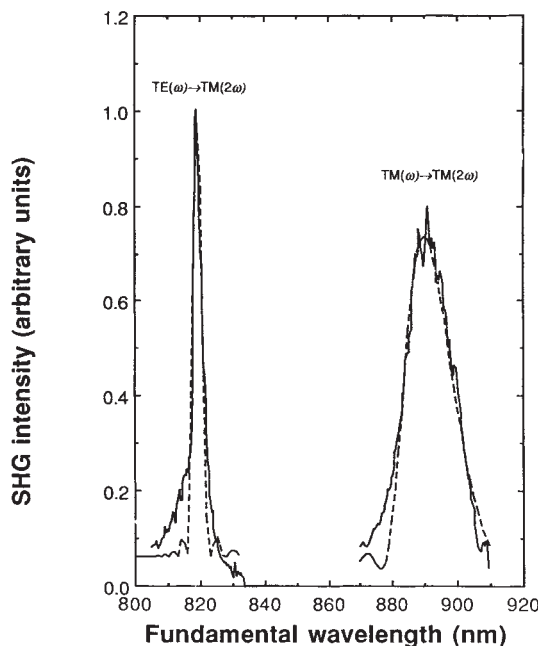


FIG. 4 Experimental measurement of the relative SHG intensity as a function of the fundamental wavelength and polarization. The solid lines are the experimental data and the dashed lines are calculated values based on the mode-conversion phase-matching scheme. SHG peaks are seen for both the predicted $\text{TE}_0(\omega) \rightarrow \text{TM}_1(2\omega)$ and $\text{TM}_0(\omega) \rightarrow \text{TM}_1(2\omega)$ mode conversions. Input polarizations were selected using a combination of a polarizer and a half-wave plate. The SHG output polarizations were confirmed by interposing a polarization analyzer. The laser source was a continuous-wave (CW) Ar⁺ ion-pumped Ti-sapphire laser tunable over 800–920 nm. The power coupled into the waveguide was measured with a power meter as ~50 mW or 5%. The fundamental beam diameter on our sample was calculated to be ~50 μm . The SHG intensities were normalized for the wavelength dependence of the square of the fundamental power and for the small difference in coupled input power between the different polarizations. The coupling angle of the input light was adjusted in 0.02° steps during the wavelength scans to maintain maximum coupling by visually keeping the m-line (the dark line in the reflected beam due to light coupled into the waveguide) centred in the reflected beam image¹⁰. A short-pass interference filter (<500 nm) was used to block the fundamental light from the photomultiplier detector. The input light was mechanically chopped at 447 Hz, and phase-sensitive detection of the calibrated photomultiplier signal allowed a detection limit of 10 pW of blue light to be achieved.

was measured directly with a power meter for 50 mW of fundamental power at 819 nm. This yields a power-corrected conversion efficiency of $0.04\% \text{ W}^{-1}$ and a normalized efficiency of $150\% \text{ W}^{-1} \text{ cm}^{-2}$, which is very high, considering that there is no lateral confinement of the light in the waveguide. For the LB film without structural inversion, a low but measurable SHG signal was detected that is smaller than in the inverted structure by a factor $>10^3$, demonstrating the dramatic enhancement of field overlap.

Although the normalized efficiency of this system is high, within a factor of four of the best electron-beam-processed lithium niobate channel waveguide¹⁷, the actual output of $\sim 1 \mu\text{W}$ is rather modest. But this system has the potential to achieve practical levels of conversion, a goal being 10% conversion for 100 mW fundamental input¹⁸. Fabrication of a stripe waveguide, for example by laser photoablation, should be able to confine the light to a diameter of $\leq 5 \mu\text{m}$, representing an order of magnitude higher power density than at present. We can also readily foresee an improvement in the nonlinear susceptibility, obtained by replacing the inert LB spacer material by an active one in which the same chromophore is linked to the polymer backbone in a reversed orientation by means of the sulphonyl group. We have already demonstrated this approach using a different chromophore, leading to a SHG enhancement of a factor of three¹⁹. These improvements would yield a second-harmonic power of 0.1 mW for 100 mW input. A longer phase-matched interaction length can lead to a large increase in performance because conversion efficiency increases quadratically with interaction length⁵. Atomic force microscopy (AFM) measurements indicate that the interaction length in the system reported here is relatively short because the inverted-structure LB waveguide deposited onto the low-index buffer has an r.m.s. roughness of 20 nm. This roughness originates in the spin-coated buffer layer, as the LB film that was simultaneously deposited directly onto the side of the substrate without the buffer layer shows a roughness of only 2–3 nm. We calculate that the conversion loss for such a smooth film due to phase mismatch would only be 10% over 1.5 mm, providing $\sim 10 \text{ mW}$ of second-harmonic light from 100 mW at the fundamental. Elimination of this buffer layer requires an increase in the refractive index of the LB film of ≥ 0.03 . Replacing the inert LB interlayer by an active one, as discussed above to improve the susceptibility, should also increase the refractive index of the waveguide beyond this threshold, making the buffer unnecessary. Thus the system has the potential for achieving a practical level of performance. $\square$

Resonant stick-slip motion in a colloidal crystal

T. Palberg & K. Streicher

Department of Physics, University of Konstanz, D-78434 Konstanz, Germany

THE frictional properties of solid bodies can often be described in terms of stick-slip motion, in which one body slides against another, at constant driving force, in an alternating series of sticking and slipping events. Stick-slip motion between surfaces separated by a very thin fluid layer was recently observed on the molecular scale¹. At the other extreme, macroscopic bodies can exhibit a coupling between stick-slip motion and mechanical (such as acoustic) resonances. The stick-slip motion of a bow on a violin string, for example, excites resonant vibration of the string². Here we describe the observation of such resonant stick-slip behaviour on the microscopic scale. We have studied the flow behaviour of colloidal crystals in a rectangular tube. These crystals, comprising a dense suspension of sub-micrometre-sized polystyrene spheres, are known to exhibit shear-induced melting at high flow rates^{3–6}. We find that, at lower flow velocity, stick-slip processes at the interface between the crystal and the cell wall can excite resonances in the crystal, detectable as periodic shifts of the Bragg angle in optical diffraction. We show that the resonances can be tuned by varying the density of the suspension, which is equivalent to altering the tension in a violin string.

Stick-slip phenomena are known to be important in many fields from friction between solids to the processing of dense suspensions or polymer melts. For two rigid solids sliding over each other with a monoatomic layer of fluid between them, highly regular stick-slip motion has been observed on the molecular scale¹. These results reflected the discrete, molecular nature of the intermediate fluid. No coupling to vibrational resonances of either solid was observed, as one might expect given the large difference in the timescales of the stick-slip periodicity and of acoustic resonances of the solids. When, on the other hand, these timescales are similar, resonances should be observable when the stick-slip motion coincides with an acoustic mode of the solid.

This is what one observes for string instruments², for which the length, diameter and tension of the strings as well as the elastic constants of the material determine the possible modes and frequencies of resonant string vibration. An arrangement of coupled resonators transforms the string vibration into a set of amplified higher harmonics. Resin powder attached to the bow regulates the adhesion between string and bow, ensuring that the stick-slip process is synchronized to the string resonances.

Resonant stick-slip motion of one body of elastic material along another is not confined to string instruments; the screech of chalk on a blackboard is another example. But a suitable model system for studying such phenomena on a microscopic scale has not previously been available, because the resonant frequencies at the atomic scale are very large (shear moduli G are of the order 10^{10} – 10^{12} N m^{-2}).

Colloidal crystals are formed easily from aqueous suspensions of sub-micrometre polystyrene spheres if the concentration of salt is lowered by contact with ion-exchange resin^{3–5}. The particles then carry a few hundred surface charges (usually dissociated carboxyl or sulphonate groups) weakly screened by the surrounding cloud of counterions (H^+). Because of mutual repulsion, the spheres arrange themselves in crystal lattices with a separation d a few times larger than their diameter and of the order of the wavelength of visible light, causing iridescence. For charged colloidal spheres the phase diagram determined by light scattering and microscopy shows regions of fluid-like, body-centred cubic (b.c.c.) and face-centred cubic (f.c.c.) order^{3–6}.

Received 24 August; accepted 10 November 1993.

1. Khanarian, G. & Norwood, R. A. in *Nonlinear Optics: Fundamentals, Materials, and Devices* (ed. Miyata, S.) 461–472 (Elsevier, Amsterdam, 1992).
2. Staring, E. G. J., Rikken, G. L. J. A., Seppen, C. J. E., Nijhuis, S. & Venhuizen, A. H. J. *Adv. Mater.* **3**, 401–403 (1991).
3. Ashwell, G. J., Hargreaves, R. C., Baldwin, C. E., Bahra, G. S. & Brown, C. R. *Nature* **357**, 393–395 (1992).
4. Roberts, G. G. *Ferroelectrics* **91**, 21–38 (1989).
5. Prasad, P. N. & Williams, D. J. *Introduction to Nonlinear Optical Effects in Molecules and Polymers* (Wiley, New York, 1991).
6. Clays, K., Armstrong, N. J., Ezenyilimba, M. C. & Penner, T. L. *Chem. Mater.* **5**, 1032–1036 (1993).
7. Penner, T. L. et al. *Molec. Cryst. Liq. Sci. Technol.* **4**, 191–209 (1993).
8. Girling, I. R. et al. *Thin Solid Films* **132**, 101–112 (1985).
9. Hönig, D. & Möbius, D. *J. phys. Chem.* **95**, 4590–4592 (1991).
10. Hunsperger, R. G. *Integrated Optics: Theory and Technology* (Springer, New York, 1982).
11. Jain, K. & Hewig, G. H. *Opt. Commun.* **36**, 483–486 (1981).
12. Williams, D. J. in *Electronic and Photonic Applications of Polymers* (eds Bowden, M. J. & Turner, S. R.) 297–330 (Am. chem. Soc., Washington DC, 1988).
13. Akhmedev, N. N. & Novak, V. R. *Opt. Spektrosk.* **58**, 558–559 (1985).
14. Stegeman, G. I., Seaton, C. T. & Zanon, R. *Thin Solid Films* **152**, 231–263 (1987).
15. Somekh, S. & Yariv, A. *Opt. Commun.* **6**, 301–303 (1972).
16. Lim, E. J., Fejer, M. M., Byer, R. L. & Kozlovsky, W. J. *Electron. Lett.* **25**, 731–732 (1989).
17. Yamada, M., Nada, N., Saitoh, M. & Watanabe, K. *Appl. Phys. Lett.* **62**, 435–436 (1993).
18. Burland, D. M., Miller, R. D., Reiser, O., Twieg, R. J. & Walsh, C. A. *J. appl. Phys.* **71**, 410–417 (1992).
19. Motschmann, H. R., Penner, T. L., Armstrong, N. J. & Ezenyilimba, M. C. *J. phys. Chem.* **95**, 3933–3936 (1993).

This phase diagram can be understood in terms of particles interacting by means of a screened Coulomb interaction with corrections for finite particle radius and a renormalized surface charge number⁵⁻⁷. For b.c.c. single crystals the (anisotropic) shear modulus G was shown to be $G = f_a n_p (\kappa d)^2 U(d)$ (refs 6, 8, 9) with n_p , κ and $U(d)$ being the particle number density, the Debye-Hückel screening parameter and the pair energy of interaction at the particle separation d , respectively; f_a is a known analytical factor related to the relative orientation of the crystal with respect to the direction of shear.

Such crystals may be excited to undergo resonant shear and

torsional vibrations if placed in resonators with typical dimensions of order a few centimetres^{8,9}. The resonant frequencies are low, typically between 1 and 100 Hz, as the shear moduli are in the range 0.1–10 N m⁻² and high frequencies are strongly damped by the suspending water. Excitation usually proceeds by means of defined periodic movement of the sample cell, with the crystal in firm contact with the cell wall.

Here we apply an external pressure difference to move a crystal along walls of an open-ended rectangular cell and exploit the stick-slip process between crystal and cell wall to excite the internal resonance. In the case of sticking to both walls, the resonant

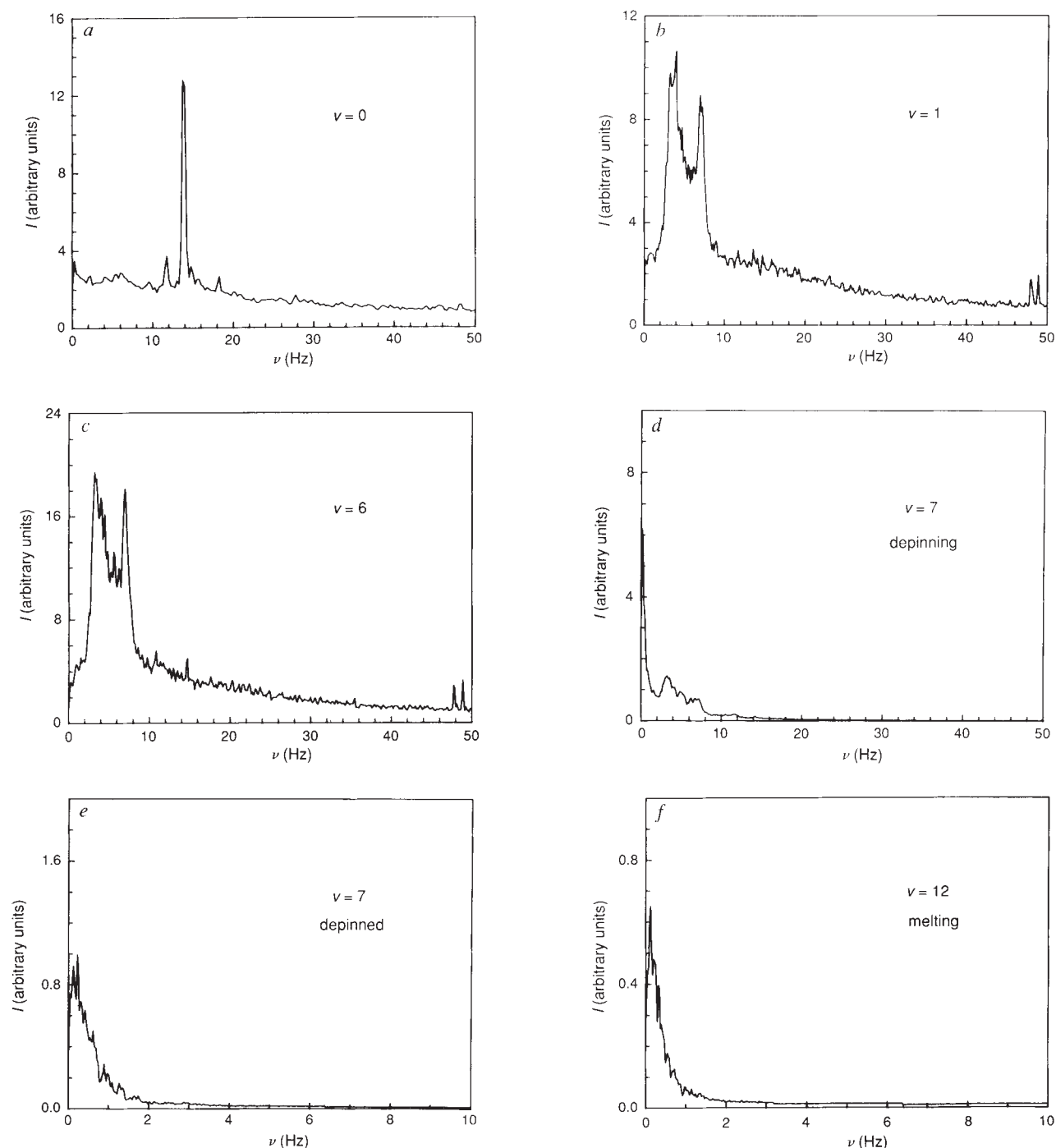


FIG. 1 Frequency spectra (intensity I against frequency ν) of a colloidal crystal *a*, at rest; *b*, *c*, excited to shear resonances by a stick-slip mechanism at relative velocities of $\nu = 1$ and $\nu = 6$, respectively; *d*, tran-

sition to continuous gliding at a relative velocity of $\nu = 7$; *e*, continuous gliding at $\nu = 7$; *f*, complete shear melting at $\nu = 12$.

frequency of this Helmholtz resonator of depth x and height y is $\omega_r^2 = G(k_x^2 + k_y^2)/\rho$, where k_x and k_y are the wave vectors of the resonant standing wave and ρ is the suspension density. For a resonator with aspect ratio $y/x \gg 1$ the contribution of k_y may be neglected. Furthermore, for a crystal sticking to only one wall the resonant frequency of the resulting $\lambda/4$ mode halves as k_x doubles. In principle, higher harmonics can also be excited, but they are strongly damped.

We used polystyrene latex spheres of hydrodynamic diameter $\sigma_h = 102$ nm at a particle concentration of $n_p = 3 \mu\text{m}^{-3}$ (Seradyn Inc., USA), the elastic properties of these spheres have been characterized extensively^{10,11}. A continuous ion-exchange procedure¹² was employed rendering the suspension completely free of small ions other than H^+ . During deionization the suspension flows through the measuring cell at velocities and shear rates well above the level for shear melting¹³. When the flow was stopped, a monolithic crystal (nucleated heterogeneously at the cell walls) formed within a few minutes. It filled the entire cell (of dimensions $x \times y \times z$, $1 \times 10 \times 40$ mm); large reservoirs at each end of the cell, however, contained a polycrystalline solid. Analysis of Laue spots appearing under illumination with a He-Ne laser showed the crystal to be b.c.c. with its (110) plane parallel to the cell wall and the [111] direction parallel to the formerly applied flow.

Shear vibrations lead to periodic shifts of the Bragg angle by way of an oscillatory change in the lattice constants of the crystal. We detect these shifts of Bragg angle by placing a position-sensitive detector at one of the Laue spots and decompose the signal into a frequency spectrum¹¹. Shear-induced phase transitions are detected by the disappearance of this (110) Laue spot.

Without flow, a low-amplitude single resonance at $\nu_0 = 14$ Hz is present as is shown in Fig. 1a. It is due to unavoidable vibrations of the tubing system corresponding to white noise. The shear modulus G is calculated to be 0.14 N m^{-2} . The slowest possible pump flow within our apparatus (relative velocity $v = 1$) is $\sim 0.01 \text{ mm}^3 \text{ s}^{-1}$ corresponding to mean velocities of $1 \mu\text{m s}^{-1}$ or one lattice constant per second. As there is no destruction of the Laue scattering pattern, and thus no irreversible distortion of the crystal lattice, we conclude the average velocity profile to be plug-like. This was confirmed for a similar sample by laser Doppler anemometry (LDA)¹⁴. Resonant phenomena are also detected during motion as can be seen in Fig. 1b and c. We observe an asymmetrically broadened peak at $\nu_2 = \nu_0/4 = 3.5$ Hz and a seemingly superimposed sharp resonance at $\nu_1 = \nu_0/2 = 7$ Hz. The frequency distribution of the ν_2 peak resembles an $1/\nu$ distribution, whereas the peak at ν_1 is symmetrical. Both frequencies are independent of velocity and stable for relative flow velocities $v < 7$. At $v = 7$, both peaks disappear while the intensity around 0 Hz drastically increases; Fig. 1d and e. This means that the still undestroyed crystal is now floating without any internal vibration. The effect is attributed to thin, shear-induced molten layers now present between the cell walls and the colloidal crystal.

The peak centred at zero frequency remains at velocities up to $v = 12$, at which speed it disappears at the same time as the Laue scattering pattern, indicating a shear-induced phase transition. The intensity of the Laue spot is shown as a function of the flow velocity in Fig. 2. The slight decrease in intensity for $8 \leq v \leq 12$ is interpreted as a successive widening of the fluid layer between the cell wall and the flowing crystal due to shear melting. After the phase transition to a fluid-like order, the typical parabolic Poiseuille flow profile is observed by LDA. A plug-like flow is not possible any longer, as the viscosity of the molten suspension now is several orders of magnitude lower¹⁴.

We will now give a simple picture of the observed colloidal crystal motion at low velocities, that is $v \leq 7$. We assume that the crystal is first pinned to both walls. As the pump is turned on, it is then slowly distorted by the applied pressure. The shear

stress is largest at the pinning areas, that is the walls. It therefore will eventually reach the yield stress, which is likely to differ slightly for the two walls, because of inhomogeneities of the wall, lattice defects and so on. Thus, after depinning at one wall, the crystal advances, relaxing into a motion according to its internal Helmholtz shear resonance. Sticking is favoured at maximum positive amplitudes, where both the relative velocity between wall and crystal becomes zero and the driving pressure difference is balanced by the elastic forces of stretching. The process is then repeated at the other wall.

As the vibrating crystal is still fixed at one wall, the resonant frequency should halve with respect to the spectrum at no flow. The peak at $\nu_1 = \nu_0/2$ therefore is identified as the mode of crystal vibration. The second asymmetrically broadened peak is present only under flow; we attribute this peak to the periodicity of the stick-slip process. Its velocity-independent frequency of $\nu_2 = \nu_0/4$ shows the stick-slip process to be independent of the grainy nature of the suspension, but strongly coupled to the shear vibration of the crystal. The crystal completes two vibrations before the maximum elongation is reached and it again sticks to the wall.

Both the creep-flow processes leading to the yielding of the colloidal crystal and the shear processes within the crystal leading to the resonant vibration are subjected to the timescale of motion of single particles, that is, the timescale of brownian motion. Therefore the asymmetry of the broadening of the stick-slip peak is readily visible in the spectrum. Systematic investigation of the frequency distribution of ν_2 should add valuable quantitative information to the discussion of the origin of noise in friction and stick-slip processes in general.

To summarize these observations on the colloidal crystal; at moderate velocities it moves in a plug-like fashion through the cell, being excited to internal acoustic resonances of velocity-independent frequencies. At higher velocities there is just a gliding movement, and at still higher velocities the crystal is completely molten. The behaviour observed for moderate velocities can be described in terms of an alternating sticking and slipping synchronized by the internal resonance of the elastic material. To our knowledge, this is the first non-macroscopic example of such a phenomenon.

Because of the fixed geometry of our sample cell, the vibrational frequency may be tuned by altering the elastic constant of the crystal, which is rather like bowing a different string on a violin. This is accomplished by varying either the particle density n_p or the salt concentration c_s in our system. Here we

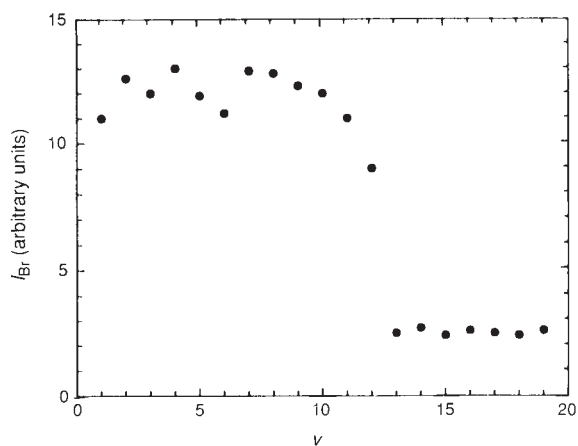


FIG. 2 Intensity of the (110) Bragg reflection (I_B) for different relative flow velocities (v). Note the slight decrease in intensity at $v \geq 8$ indicating successive widening of the fluid layers between wall and crystal, and the sudden drop of intensity at $v = 12$, indicating complete shear melting.

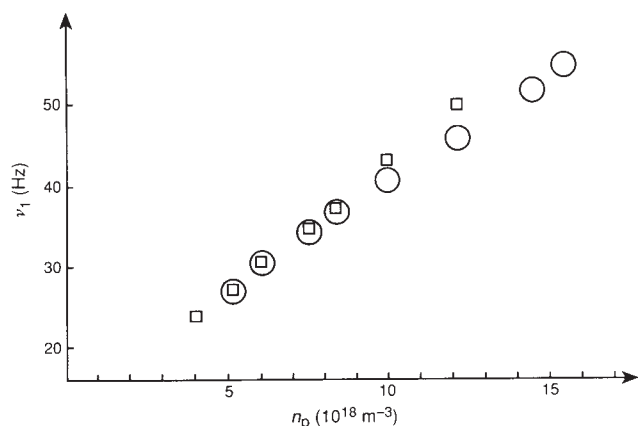


FIG. 3 Resonant frequencies ν_1 as a function of the particle number densities n_p for a polycrystalline colloidal solid. Squares show experimental data; circles indicate calculated values of n_p needed at completely deionized conditions to generate the frequencies of one octave (27.5–55 Hz).

measured the effect on frequency of varying n_p for a polycrystalline sample that was grown after the addition of a sub-micromolar quantity of NaOH. This suppresses heterogeneous nucleation but does not alter the shear modulus⁹. In Fig. 3 we compare the n_p dependence of measured ν_1 frequencies to theoretical values ($f_a = 2/9$). The agreement is very good at low n_p . The observed systematic deviation at higher particle densities is due to the decreasing size of individual crystallites, as will be discussed elsewhere.

Although there are obvious differences in material and geometrical properties, the colloidal system presented here constitutes a mesoscopic analogue of a string excited by the movement of a bow. In both cases the frequency of the exciting stick-slip mechanism is synchronized to the tunable vibrational resonance of the elastic material. The energy transfer into vibrations is severely limited, however, as the crystal is detached from the cell walls at higher velocities. Further, there is no additional resonance corpus present to transform the energy of the crystal vibration into perceptible sound waves. For these reasons, the colloidal crystal violin (or double bass) will probably not play at the Royal Albert Hall. Nevertheless it may well establish its place in solid-state physics for its unique feature of resonant stick-slip behaviour that is easily observable on a mesoscopic scale. □

Direct anthropogenic contributions to sea level rise in the twentieth century

Dork L. Sahagian*, Frank W. Schwartz* & David K. Jacobs†

* Department of Geological Sciences, Ohio State University, Columbus, Ohio 43210, USA

† Department of Invertebrates, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA

GLOBAL compilations of tide records indicate that sea level has been rising throughout the twentieth century^{1,2}, with potentially dangerous consequences for low-lying coastal regions. Thermal expansion of ocean water³ and melting of alpine glaciers⁴ in response to increased atmospheric temperatures may have been responsible for some of this change, but human activities can also influence sea level directly. For example, the rise in sea level would have been even larger⁵ if large quantities of water had not been stored in reservoirs, and channelled into aquifers by irrigation projects. Here we show, however, that these and other human activities have together caused a net increase in sea levels over the past century. We estimate that a combination of groundwater withdrawal, surface water diversion and land-use changes has caused at least a third of the observed rise, and suggest that the contributions of climate-related effects must therefore be smaller than has been previously supposed.

Natural variations in surface- and groundwater storage⁶ caused by past climate changes are thought to have produced rapid low-amplitude variations in sea level on geological timescales⁷. Human activities can directly influence sea level as well. Natural groundwater systems typically are in a condition of dynamic equilibrium where, over long times, recharge and discharge are in balance. When the rate of groundwater pumping greatly exceeds the rate of recharge, as is often the case in arid or even semi-arid regions, water is removed permanently from storage. Ground water lost from storage in this way is termed 'mined water'. Eventually, depletion must stop once pumping is prohibitively expensive, or the reservoir is depleted. In arid areas that are being irrigated by ground water, the cycling of water may be complex. For example, some of the ground water pumped to the surface may return to the aquifer when irrigation rates exceed crop utilization rates, or through leakage from surface impoundments. On balance however, the large evaporative loss inherent with irrigation and surface storage results in a net, local transfer of water out of the groundwater portion of the hydrologic cycle. The water that is lost from groundwater storage eventually reaches the ocean through the atmosphere or surface flow, and ultimately results in sea level rise. We examine the impact on sea level from mining of aquifers in North America, North Africa and Arabia, consider lake and groundwater loss from the Aral and Caspian basins, and estimate moisture losses due to desertification of the Sahel, and the destruction of forests and wetlands (Table 1). We do not include smaller aquifers or other reservoirs nor do we extrapolate our results beyond the regions where data are available. Consequently, our results provide a minimum estimate of direct anthropogenic contributions to sea level rise. The minor issue of isostatic adjustment of sea floor or continents to water mass redistribution is not considered.

There are three large aquifers in North America associated with limited recharge and significant mining. The High Plains regional aquifer includes the Ogallala formation and extends from Texas to South Dakota. Although the High Plains receive significant rainfall, most of this is returned to the atmosphere by evapotranspiration⁸, and so recharge rates are small (0–41 mm yr⁻¹) (ref. 9). Recharge rates are generally difficult to

Received 28 July; accepted 8 November 1993.

1. Gee, M. L., McGuiggan, P. M., Israelachvili, J. N. & Homola, A. M. *J. chem. Phys.* **93**, 1895–1906 (1990).
2. Fletcher, N. H. & Rossing, T. D. *The Physics of Musical Instruments* (Springer, New York, 1992).
3. Pusey, P. N. in *Liquids, Freezing and the Glass Transition* (eds Hansen, J. P., Levesque, D. & Zinn-Justin, J.) 763–942 (North Holland, Amsterdam, 1991).
4. Ackerson, B. J. *Phase Transitions* **21**, 2–4 (1991).
5. Voegtli, L. P. & Zukoski, C. F. *J. Coll. Interface Sci.* **113**, 79–91 (1991).
6. Robbins, M. O., Kremer, K. & Grest, G. J. *chem. Phys.* **88**, 3286–3321 (1988).
7. Alexander, S. et al. *J. chem. Phys.* **80**, 5776–5781 (1984).
8. Dubois-Violette, E., Pieranski, P., Rothen, F. & Strzelecki, L. *J. Physique* **41**, 369–376 (1980).
9. Chaikin, P. M., diMeglio, J. M., Dozier, W., Lindsay, H. M. & Weitz, D. A. in *Physics of Complex and Supermolecular Fluids* (eds Safran, S. & Clark, N. A.) 65–81 (Wiley, New York, 1987).
10. Palberg, T. et al. *J. Coll. Interface Sci.* (submitted).
11. Palberg, T. et al. *J. Phys.* (submitted).
12. Palberg, T. et al. *J. phys. Chem.* **96**, 8180–8184 (1992).
13. Ackerson, B. J. & Clark, N. A. *Physica* **118A**, 221–249 (1983).
14. Palberg, T., Würth, M., König, P., Simnacher, E. & Leiderer, P. *Prog. Colloid Polym. Sci.* **89**, 125–126 (1992).
15. Chen, L. B. & Zukoski, C. F. *Phys. Rev. Lett.* **65**, 44–47 (1990).

ACKNOWLEDGEMENTS. This work was supported by the Deutsche Forschungsgemeinschaft.

TABLE 1 Anthropogenic contributions to sea level rise

Water reservoir	Total volume removable ($\times 10^{12} \text{ m}^3$)	Sea level equivalent (cm)	Present net extraction rate ($\times 10^{10} \text{ m}^3 \text{ yr}^{-1}$)	Sea level rise rate (mm yr^{-1})	Projected sea level change next 50 yr (mm)	Estimated sea level change to date (mm)
High Plains	4.0	1.1	1.2	0.03	1.6	1.1
SW US	3.0	0.83	1.0	0.03	1.5	0.92
California	10.0	2.7	1.3	0.04	1.9	1.2
Sahara	600	167.0	1.0	0.03	1.4	0.56
Arabia	500	140.0	1.6	0.04	2.2	0.89
Aral (lake)						
1960	1.1	0.3	2.7	0.08	3.0	2.2
1990	0.3	0.08				
Aral (ground water)	2.2	0.6	3.7	0.1	5.1	3.1
Caspian (lake)	56.0	15.4	0.77	0.02	1.1	1.3
Caspian (grdwtr)	220.0	61.2	0.47	0.01	0.65	0.78
Sahel (soil water)	0.1	0.03	0.34	0.01	0.5	0.28
Deforestation	3.3	0.9	4.9	0.14	6.8	3.4
Wetland reduction	8.6	2.4	0.2	0.006	0.3	1.3
Dams	-1.9	-0.52	—	—	—	-5.2
Total	1,406.7	392.1	19.2	0.54	26.1	11.8

'Total volume removable' is the amount of water that could be withdrawn economically from each aquifer with present-day technology. Methods of volume estimation vary from one aquifer (and economic condition) to another, so these figures should be considered as rough approximations. 'Sea level equivalent' is the amount that eustatic sea level would rise if the total removable volume were added to the oceans with an area of $3.6 \times 10^{14} \text{ m}^2$. 'Present net extraction rate' is the rate of water removal after accounting for recharge rates of 6.75×10^9 , 1.05×10^9 and $7.8 \times 10^8 \text{ m}^3 \text{ yr}^{-1}$ for the High Plains, Southwest US and California aquifers, respectively. Saharan and Arabian aquifers are not recharging significantly. 'Projected sea level change' is the volume of water that would be withdrawn in 50 yr if the extraction rate remained the same as it is at present. 'Sea level change to date' is the contribution of each aquifer to the twentieth-century sea level rise. For most cases, this was calculated by assuming a linear increase in extraction rate from 0 to the present rate. The increase is assumed to have begun in 1930 for American aquifers, 1950 for Saharan and Arabian aquifers, 1960 for Sahel desertification and 1940 for deforestation. Deforestation figures include only losses of tropical forests. Wetland reduction includes the total global wetland area in total removable volume, but the rates and projections include only the loss of wetlands in the United States. If the rate of reduction in the rest of the world is equal to the US rate, these figures should be doubled, and thus represent a very conservative estimate. The Aral and Caspian histories are well recorded by their fluctuating levels. We assume a specific yield of 0.2 in the sands in the surrounding desert, over an area of ~ 5 times the area of the lake itself for the Aral, and 3 times the area for the Caspian. At the present rate of reduction, the Aral will be gone in ~ 20 yr. The negative values for dammed reservoirs reflect excess water storage on land, assuming all reservoirs are constantly filled to capacity. The river damming programs of the 1960s and early 1970s have tapered off.

calculate *a priori*, but the significant water table depression indicates that withdrawals of ground water overwhelm what little recharge there is. There are $\sim 200,000$ wells that extract groundwater from the High Plains regional aquifer. The total pumping rate has been increasing since about 1930. The southwest US aquifer¹⁰ encompasses the alluvial basins of southern New Mexico, Arizona, California and northern Sonora. The aquifer of the Central Valley of California is developed by wells to a variable extent and exhibits broad variability in its geology and hydrogeology. In the northernmost part of the valley, recharge may be keeping up with withdrawal rates. In the central and southern parts of the valley, however, recharge is minimal and a substantial quantity water has been mined, resulting in sediment compaction and subsidence¹¹. Total recoverable ground water in these American aquifers is equivalent to ~ 4.6 cm in sea level, is presently being mined at a sea-level-equivalent rate of 0.1 mm yr^{-1} , and has contributed to 3.2 mm of sea level rise to date (Table 1). The future of groundwater mining in North America will depend on the complex interplay between economic and political considerations of the allocation of surface and groundwater between separate basins and states. Assuming continued mining at present-day rates for the next 50 yr, another 5 mm increase in sea level would result from these aquifers alone.

North Africa includes several sedimentary basins that contain large non-recharging aquifers¹². Although the basins in North Africa contain a large volume of water, it is not likely that all of the pumpable water will ever be extracted because there is no extensive irrigation in the region. About $10^{10} \text{ m}^3 \text{ yr}^{-1}$ of water are being withdrawn from the northern Sahara¹³. The total volume of ground water in Africa has been estimated as $6 \times 10^{15} \text{ m}^3$ (ref. 14). But only one quarter of this occurs in non-recharging aquifers, which implies that the potential for groundwater mining throughout Africa is $\sim 1.5 \times 10^{15} \text{ m}^3$, a sea-level-equivalent of ~ 4 m. In our calculations, we include only the contribution

of northern Saharan groundwater mining.

Arabia is underlain by an aquifer up to 3 km thick in many places. The volume of ground water stored in the aquifer system is $5 \times 10^{14} \text{ m}^3$ (ref. 15), equivalent to 1.4 m of sea level. In the years 1985–90, $\sim 2.6 \times 10^{10} \text{ m}^3$ of water were withdrawn from this aquifer, and production is rapidly increasing. At a constant 1987 rate of $1.6 \times 10^{10} \text{ m}^3 \text{ yr}^{-1}$ (ref. 16), the loss of ground water from storage by 2010 will be $\sim 3.2 \times 10^{11} \text{ m}^3$, or a sea-level-equivalent of 0.9 mm. Given the vast groundwater reserves and the increasing demand for irrigation in the region, groundwater extraction could increase dramatically.

Diversion of surface waters for irrigation in the internally draining basins of arid regions results in increased evaporation and net loss of water from the basin hydrologic system. The Aral Sea is fed by the Amu Dar'ya and Syr Dar'ya rivers, which have been diverted for cultivation of rice and cotton resulting in increased evaporative loss and reduced river flow¹⁷. The level of the Aral Sea has varied through recorded history in part as a consequence these diversions, and it has lost 16 m in depth and half its surface area ($3.3 \times 10^{10} \text{ m}^2$) between 1960 and 1990 (ref. 18). Former coastal cities are now many kilometres from the receding shoreline, and lake volume is declining at a rate of $2.7 \times 10^{10} \text{ m}^3 \text{ yr}^{-1}$. Assuming that groundwater drainage has kept pace with lake level reduction, it has contributed $3.7 \times 10^{10} \text{ m}^3 \text{ yr}^{-1}$ for a total flux out of the Aral basin of $6.4 \times 10^{10} \text{ m}^3 \text{ yr}^{-1}$.

The Caspian Sea is fed primarily by the Volga River, which drains much of European CIS. Water levels fell 3 m from 1929 to 1977, then rose 1.75 m by 1990 (refs 19, 20). The causes of these Caspian fluctuations are poorly understood. Despite diversion of water from the Volga, the level is presently rising, presumably as a consequence of climatic variations in the Caspian basin and Volga drainage basin^{20,21}. Thus, the Caspian is responsible for a net 1.3 mm of eustatic sea level rise this

century, but the trend of its water contribution to the ocean has reversed. The groundwater contribution was calculated assuming a specific yield of 0.2 over a basin area of three times the lake water area.

Human activity affects the water storage capacity of soils, forests and wetlands. In the Sahel, desertification has reduced soil moisture from ~2% to virtually zero, partly as a result of changing agricultural and grazing practices²². If a 5-m-thick soil has been affected over an area of $1 \times 10^{12} \text{ m}^2$ since 1960, then the Sahel has contributed 0.28 mm to sea level rise.

Forests store water both in living tissue and in the high soil moisture maintained by plant cover. If an average tropical forest has a biomass of 45 kg m^{-2} (ref. 23), and the mass is converted to water by conversion of carbohydrates (cellulose, $(\text{C}_6\text{H}_{10}\text{O}_5)_n$) to CO_2 and H_2O , it would be equivalent to 2.5 cm (25 litres m^{-2}) of water. If we assume a bulk forest dry-to-wet mass ratio of 0.25, then an additional 13.5 cm is stored in plants (biomass is defined as dry matter²³). We estimate an equal amount of water to be stored in soil, fallen vegetation, on leaves and in the surrounding atmosphere, making a total of 36.0 cm of water that would be released by the destruction of a forest. Deforestation of the tropical regions alone at a present rate of $1.5 \times 10^{11} \text{ m}^2 \text{ yr}^{-1}$ (ref. 24) would add $4.9 \times 10^{10} \text{ m}^3 \text{ yr}^{-1}$ water to the oceans, causing a sea level rise of 0.14 mm yr^{-1} , or 6.8 mm over 50 yr. This value is larger than the depletion rate of any other continental water storage reservoir, yet does not include the rate of deforestation of temperate and northern woodlands.

In the United States, wetlands have been drained at a rate of $2.2 \times 10^9 \text{ m}^2 \text{ yr}^{-1}$ since 1780 (refs 25, 26). Wetlands contain standing water, soil moisture, and water in plants, equivalent to water ~1 m deep²⁷. Global wetland area ($8.56 \times 10^{12} \text{ m}^2$ (ref. 25)) is much larger than that in the United States, so the total contribution of wetland destruction to sea level rise is likely to be larger than we have calculated.

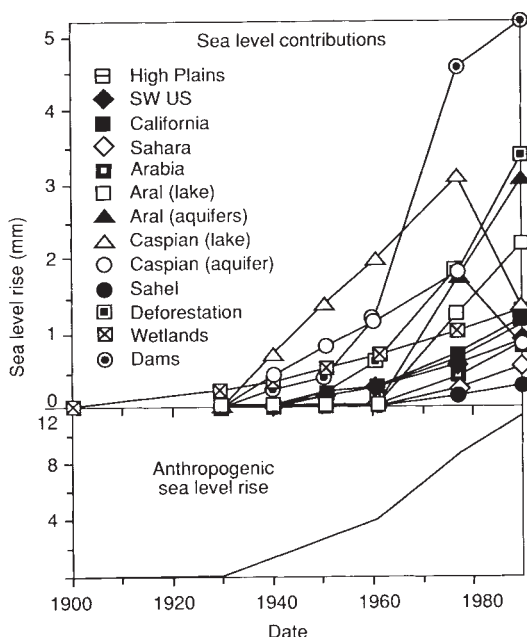


FIG. 1 Anthropogenic contributions to twentieth-century sea level rise. Upper part indicates individual contributions of various sources. Lower part is total anthropogenic sea-level rise history. The plots of most contributions assume a linear increase in withdrawal rate throughout the twentieth century. Note that dams serve to counteract sea level rise by storing additional water on the continents, and that the 'dams' curve should be considered as having negative values. Wetlands are plotted assuming all wetland destruction in the United States has occurred after 1900.

Our calculations include only the largest and best documented water sources for each of the categories we have considered. We have not considered many small aquifers or the large Australian aquifers which are currently mined at very low rates. considerable water and oil are extracted at oil wells. This would have an effect on sea level to the extent that a larger volume of fluid is extracted than water injected to enhance recovery. Polders (dikes) artificially maintain areas below sea level free of water and thus make a small contribution to sea level rise, but are not considered in this analysis. Human activities may have increased the rate of stream down-cutting in the twentieth century, resulting in removal of ground water in the stream drainage area. Sediments eroded from (cultivated) land are deposited in the ocean and displace sea water. We have not calculated the magnitudes of these effects, but in combination with other factors, they may be significant. Thus, the totals in Table 1 should be considered minimum estimates.

Some human activities may store additional water on the continents rather than supplying it to the ocean²⁸. The volumes of large dammed lakes³⁰ (a total of $\sim 1.9 \times 10^{12} \text{ m}^3$ or equivalent to 5.2 mm of sea level) should be subtracted from the total anthropogenic sea level rise.

The total anthropogenic sea level rise resulting from the sources we have considered is 11.8 mm, with an average rate of rise since 1960 of 0.54 mm yr^{-1} (considerably greater than the value cited by Meier²⁹). The anthropogenic sea level curve (Fig. 1) can be compared to measured twentieth-century sea level variation inferred from tide gauge records¹. Peltier and Tushingham¹ have applied models of lithospheric flexure and mantle rheology to estimate and eliminate the effect of post-glacial rebound, and have obtained a net signal of $2.4 \pm 0.9 \text{ mm yr}^{-1}$ of sea level rise for the twentieth century. Using a similar approach, Trupin and Wahr³¹ obtained a rate of sea level rise of $1.5 \pm 0.4 \text{ mm yr}^{-1}$, with a preferred value of 1.75 mm yr^{-1} , which is consistent with Peltier and Tushingham¹. These rates are slightly greater than values obtained earlier (1 mm yr^{-1}) by Gornitz *et al.*², and Fairbridge and Krebs³². If a figure of 1.75 mm yr^{-1} is used for comparison, then human activities directly contribute at least 30% of the twentieth-century sea level rise. Consequently, the contributions of glacial melting and ocean thermal expansion are smaller than previously thought. It should be noted that deforestation is providing a greater contribution to current sea level rise than any other single source. Increased rates of groundwater mining from the extremely large aquifers in Arabia and Africa could, however, have important implications for sea level rise in the twenty-first century. There are additional water sources not considered in this preliminary analysis, so the rate of $1.9 \times 10^{11} \text{ m}^3 \text{ yr}^{-1}$, or 0.54 mm yr^{-1} of sea level rise should be taken as a minimum estimate. On the basis of our calculations, it seems that further study and continued monitoring of the flux of water in and out of continental reservoirs is warranted. □

Received 12 July; accepted 10 November 1993.

- Peltier, W. R. & Tushingham, A. M. *J. geophys. Res.* **96**, 6779–6804 (1991).
- Gornitz, V. & Lebedeff, S. in *Sea-Level Fluctuation and Coastal Evolution* (eds Nummedal, D., Pilkey, O. H. & Howard, J. D.) 3–16 (Soc. Econ. Paleont. Miner, Tulsa, 1987).
- Roemmich, D. *Science* **257**, 373–375 (1992).
- Meier, M. F. *Science* **226**, 1418–1421 (1984).
- Newman, W. S. & Fairbridge, R. W. *Nature* **320**, 319–321 (1986).
- Hay, W. & Leslie, M. in *Sea Level Change* (ed. Revelle, R.) 161–170 (Geophys. Study Committee, National Research Council, Washington DC, 1990).
- Jacobs, D. K. & Sahagian, D. L. *Nature* **361**, 710–712 (1993).
- Weeks, J. & Gutentag, E. in *Hydrogeology* (eds Back, W., Rosenshein, J. & Seaber, P.) 157–164 (Geol. Soc. Am., Boulder, 1988).
- Nativ, R. *Envir. Geol. Water Sci.* **19**, 21–32 (1992).
- Anderson, T., Welder, G., Lesser, G. & Trujillo, A. in *Hydrogeology* (eds Back, W., Rosenshein, J. & Seaber, P.) 81–86 (Geol. Soc. Am., Boulder, 1988).
- Farrar, C. & Bertoldi, G. in *Hydrogeology* (eds Back, W., Rosenshein, J. & Seaber, P.) 59–68 (Geol. Soc. Am., Boulder, 1988).
- Balek, J. *Hydrology and Water Resources in Tropical Africa* 1–208 (Elsevier, New York, 1977).
- Kehinde, M. & Loehnert, E. *J. Afr. Earth Sci.* **9**, 179–185 (1989).
- Castany, G. *Episodes* **15**, 18–22 (1981).
- Al-Ibrahim, A. A. *Ambio* **20**, 34–37 (1991).
- Al-Saleh, M. A. *Geogr. J.* **158**, 215–222 (1992).

17. Kotlyakov, V. M. *Environment* **33**, 4–9 (1991).
18. Micklin, P. *Post-Sov. Geogr.* **33**, 269–282 (1992).
19. Bond, A. R. *Sov. Geogr.* **32**, 516–518 (1991).
20. Rodionov, S. N. *Sov. Geogr.* **31**, 265–275 (1990).
21. Golitsyn, G., Dzyuba, A., Osipov, A. & Panin, G. *Dokl. Akad. Nauk SSSR* **313**, 1224–1227 (1990).
22. Lo, H. & Sene, A. *Int. Soc. Sci. J.* **41**, 449–456 (1989).
23. Rohrig, E. in *Ecosystems of the world* (ed. Ulrich, E.) 165–174 (Elsevier, New York, 1991).
24. Aldous, P. *Science* **259**, 1390 (1993).
25. Mitsch, W. & Gosselink, J. *Wetlands* (Van Nostrand Reinhold, New York, 1986).
26. Mitchell, G. J. in *Wetland Creation and Restoration: The Status of the Science* (eds Kusler, J. & Kentula, M.) ix–x (Island, Washington DC, 1990).
27. Wingham, D., McCormick, J., Good, R. & Simpson, R. in *Freshwater Wetlands: Ecological Processes and Management Potential* (eds Good, R., Wingham, D. & Simpson, R.) 3–20 (Academic, New York, 1978).
28. Newman, W. & Fairbridge, R. *Nature* **320**, 319–321 (1986).
29. Meier, M. *Nature* **343**, 115–116 (1990).
30. Tilsley, J. *Major Dams, Reservoirs, and Hydroelectric Plants—Worldwide* (US Dept of Interior, Bureau of Reclamation, Denver, 1983).
31. Trupin, A. & Wahr, J. *Geophys. J. Int.* **100**, 441–453 (1990).
32. Fairbridge, R. & Krebs, O. *Geophys. J. R. astr. Soc.* **6**, 532–545 (1962).

Overprinting of natural magnetic remanence in lake sediments by a subsequent high-intensity field

Robert S. Coe* & Joseph C. Liddicoat†

* Earth Sciences Department, University of California, Santa Cruz, California 95064, USA

† Environmental Sciences, Barnard College, Columbia University, 3009 Broadway, New York, New York 10027-6598, USA

THERE has been considerable debate recently on how well sedimentary rocks record transitional field directions during a polarity reversal^{1–5}. This has not been an easy question to answer because the input geomagnetic signal for a given sedimentary record is not known independently, and because the process by which remanent magnetization is acquired by sediments is difficult to duplicate in the laboratory. The geomagnetic excursion recorded by sediments outcropping around the shores of Mono Lake, California^{6–8}, offers an unusually good opportunity to examine this question. We report here a study of the Mono Lake Excursion (MLE) at a new locality on the southeast shore where, although the records satisfy usually applied standards of palaeomagnetic quality, it is clear from comparison with records from other localities that the field directions during a time of low field intensity have not been faithfully preserved. The distinctive directions of natural remanence in the discrepant part of the record show that the southeast shore sediments were overprinted by the ensuing higher-intensity field acting on the still unconsolidated sediment, apparently by realigning some magnetic grains that had previously been locked in.

Mono Basin (Fig. 1) has proved to be an almost ideal natural laboratory in which to study the fidelity of sedimentary magnetization. Well-exposed sections outcrop in several places around the lake and contain numerous tephra marker horizons that enable unambiguous identification of stratigraphic position. The lake sediments are mainly fine-grained silt composed of glacial flour derived from granitic and metamorphic rocks of the Sierra Nevada, with smaller amounts of volcanic material and clastic carbonate⁹. Bedding is regular and only rarely disturbed, the deposition rate based on ¹⁴C dates¹⁰ is high (25 to 40 cm kyr⁻¹), and a stable component of natural remanent magnetization (NRM) that decays univectorially to the origin is obtained easily by either alternating field (AF) or thermal demagnetization techniques^{7,8}. Finally, previous studies⁸ of the MLE in seven sections at three localities around the lake have obtained highly reproducible recordings using AF demagnetization (Fig. 1) that can serve as a standard signal against which to compare other

recordings of the excursion. Indeed, a longer record¹⁰ at one of these localities, Wilson Creek (Fig. 1), that includes the MLE has been used successfully as a reference for temporal correlation of lake sediments in the western United States¹¹.

The MLE is precisely located by a distinctive tephra layer (Wilson Creek Ash Bed 15, referred to as Ash 15), which divides it into two parts of roughly equal duration. Interpolation between ¹⁴C-dated horizons (assuming constant sedimentation rate) suggests the MLE began 29,000 ¹⁴C-yr ago and lasted 2,000 yr (refs 7, 8, 10). (Note that various studies^{12,13} indicate that ¹⁴C ages for this time are several thousand years too young.) The magnetic signature is distinctive (Fig. 1): a sharp swing in direction to the west and up at lower NRM intensity is followed by a swing to the east and steeply down at higher NRM intensity. The NRM and partially demagnetized NRM vary similarly with depth in all seven sections, consistent with our earlier relative palaeointensity experiments based on NRM/ARM ratios at 20 mT demagnetization level (see Table 3 and Fig. 9a in ref. 7). These results indicate that the field was as much as four times weaker during the first, older directional swing of the excursion than during the second, younger swing.

At a new locality (Fig. 1) in wave-cut bluffs on the southeastern shore of Mono Lake, however, the signature of the MLE obtained by AF demagnetization of samples from three parallel sections separated by 10–600 m is significantly different (Fig. 2a) from that found on the northwestern and eastern side of the lake. In the first part of the excursion, below the Ash 15 marker, the swing to negative inclination and westerly declination is absent. But the following swing to steeply downward directions in the second part of the excursion signature is reasonably well represented, bearing in mind that the large differences in declination in Fig. 2a stem from much smaller differences in overall direction due to the high inclinations (87–60°).

What could be the explanation of the discrepancy in magnetic signatures? The palaeomagnetic directional records strongly suggest that the sediments deposited at the southeastern shore during the first part of the MLE bear a partial imprint of the succeeding part of the excursion. The evidence for this is contained in the 30 cm of section just below Ash 15 (Fig. 2a), which are magnetized with unusual directions that are much more easterly and substantially steeper (average N50° E, 70° down) than either the axial or today's inclined dipole field. Thus, conventional viscous overprinting—the continual drift of remanent magnetization towards equilibrium with the *ambient* field—is ruled out. What field, then, could have produced an overprint with this distinctive direction? The obvious candidate is the steep easterly field of the second, higher-intensity part of the excursion.

A number of alternative hypotheses for the absence of the first directional swing in the three southeastern shore records do not stand up to scrutiny. There is no evidence for local disturbances such as slumping or for a hiatus in any of them. It is not probable that we missed the earlier swing by failing to sample low enough in the section, because a several hundred per cent jump in sedimentation rate would be required to push the missing swing below our lowest samples. Moreover, with either of the above hypotheses we would still be left without an explanation for the unusually steep and easterly directions that we find in place of the swing. Finally, to suppose instead that the magnetic signature of the southeast shore is the more accurate representation of the geomagnetic excursion is unrealistic for several reasons. The agreement of the records carrying the usual MLE signature (Fig. 1) is extraordinary by palaeomagnetic standards, distinctly better than the still moderately good agreement of the three records from the southeastern shore (Fig. 2a). Although reproducibility of results between sections is not sufficient to prove accuracy, it is a necessary criterion and is far and away the most commonly employed test for reliability of palaeomagnetic recording. An even more telling argument, in our opinion, is the lack of a plausible explanation other than magnetization during a geomagnetic field excursion for the

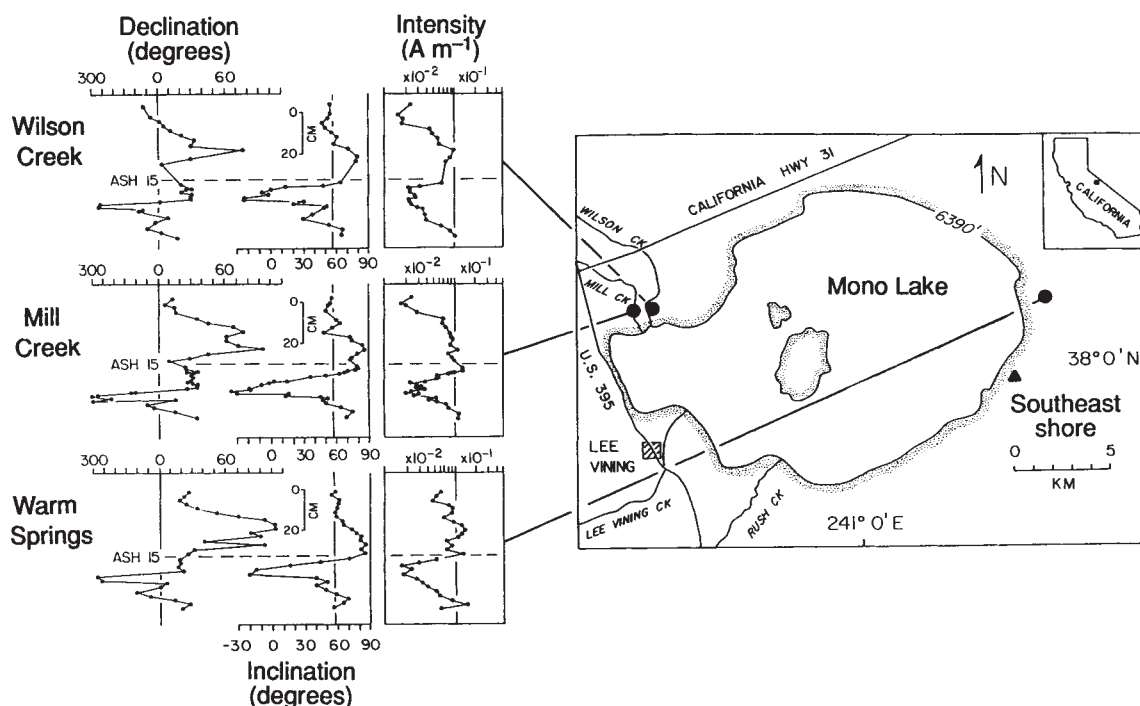


FIG. 1 Localities and corresponding palaeomagnetic records from previous studies (filled circles of the Mono Lake Excursion (MLE)^{6-8,15}). These highly reproducible records define the standard signature against which we compare other recordings of the MLE. The location of this

new study at the southeastern shore is indicated by a filled triangle, representing three sections separated by 10–600 m (see Fig. 2 for corresponding records).

highly reproducible standard records of the first swing. The recent suggestion by Quidelleur and Valet (personal communication, 1993) that spuriously shallow inclinations may result from the greater relative influence of gravity when the magnetic field is weak cannot apply in this case because the directions swing through the horizontal plane to 30° negative inclination. In contrast, assuming that the standard records from the other localities *do* represent the actual geomagnetic field, the discrepant directions found at the southeastern shore below Ash 15 can be explained quite naturally: they are the result of overprinting by the stronger, steep and easterly field that followed during the second part of the excursion.

What kind of process could produce the observed overprinting? Not simple time averaging of the MLE signal; this process could smooth away the first directional swing (if we assume that there had been an unusually thick remanence lock-in zone during deposition of the southeastern-shore sediments) but cannot account for the lack of smoothing of the second directional swing relative to the standard signal above Ash 15 (Fig. 2a). Rather, some previously locked magnetic grains below Ash 15 must have been realigned, the stronger field during the second part of the MLE supplying the extra torque needed to overcome the frictional restraining forces on some of the magnetic grains in the still wet and unconsolidated sediments. This is the same mechanism proposed very recently by Lovlie¹⁴, who simulated it by inducing remanence in suspensions of magnetite grains in slowly curing epoxy resin. The essential difference from the simpler time-averaging processes commonly discussed with regard to post-depositional remanent magnetization is that the lock-in depth of a magnetic grain below the water-sediment interface is not constant; instead, because the mechanical constraints restricting grain realignment increase continuously during consolidation, lock-in depth increases with field strength. With the process suggested here, the later strong-field swing to steep inclinations would be preserved because the ensuing field was weaker.

Why the sediments exposed along the southeastern shore were susceptible to this kind of overprinting, and the sediments from

the other localities around the lake are not, is a good question that we cannot answer yet. Because the sediments are more finely laminated and were deposited at a 60% higher rate at the south-east shore, we thought originally that they would carry an even higher-resolution record of the MLE. A summary of relevant information is as follows. X-ray diffraction and Curie-point determinations on magnetic concentrates of sediment sampled at Wilson Creek indicate only one magnetic mineral, titanium-poor magnetite (Fig. 2a in ref. 7; ref. 15). Thermal demagnetization of samples from all sections removes 95–99% of the NRM by 575 °C, consistent with the identification of magnetite. Agreement between stable remanence directions from thermal and alternating field demagnetization is extremely good right through the excursion, as shown for paired specimens from the same samples (taken by hand) from the southeastern shore (Fig. 2b) and Mill Creek (not shown). Thus Mono Lake sediments do not suffer from the same problem that was reported for Lake Tecopa sediments¹⁶, in which paired samples through the Brunhes–Matuyama transition were found to give very different results under AF and thermal demagnetization. Anisotropy of magnetic susceptibility measurements of remaining undemagnetized samples above Ash 15 at the southeastern shore and Mill Creek localities reveal typical results for undisturbed sedimentary fabrics, with minimum axes within 5–10° of the normal to bedding, 10% foliation and 1% lineation. We conclude that a more thorough magnetic and sedimentological characterisation is needed to shed light on the essential differences in properties between the southeastern shore section and the other sections. This will require resampling the critical sections to obtain suitable material for study.

Records of the MLE have been reported from four other localities in western North America, and at only one of them is the earlier directional swing well recorded (Gulf of California¹⁷). At the other three it is missing or only vaguely expressed (Summer Lake in Oregon^{11,18} and Carson Sink and Pyramid Lake in Nevada⁸). These earlier records did not drive us to the conclusion that the low-field part of the MLE was overprinted

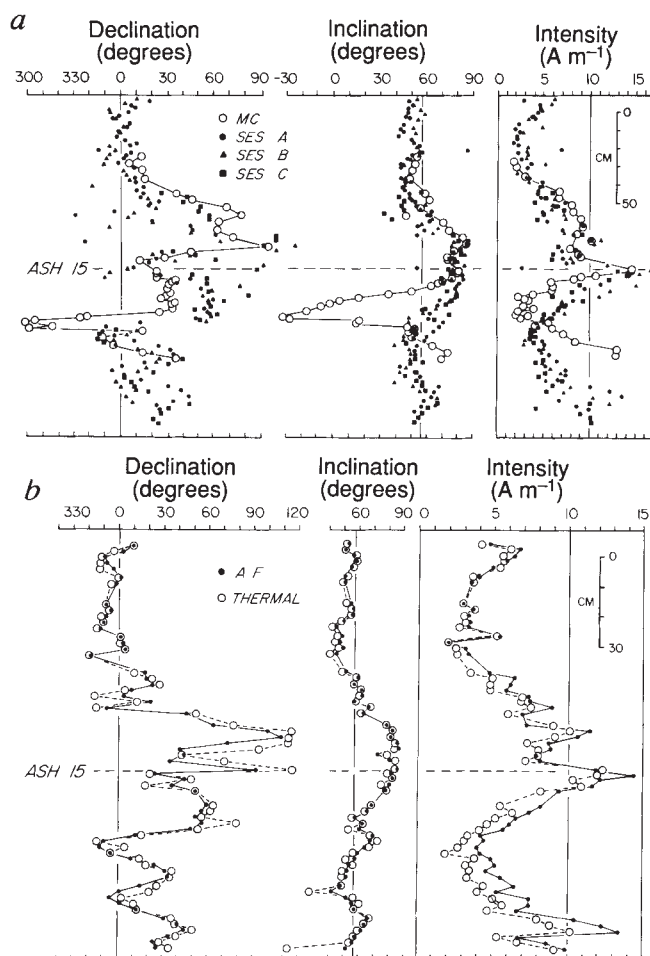


FIG. 2 a, Comparison of the three new recordings (filled symbols, SES, A, B, C) of the Mono Lake Excursion obtained from sediments at the southeastern shore against the standard MLE signal (open circles) from Mill Creek locality (see map in Fig. 1). Note that the first directional swing, below Ash 15, is missing in the southeastern shore records. A common depth scale (shown vertical) is achieved by stretching the MC record by 60%, so that the interval between Ash 15 and overlying Ash 14 is the same as at the southeastern shore (0.7 m). b, Comparison of AF (filled circles) and thermal (open circles) demagnetization results for adjacent pairs of specimens from the southeastern shore sampled through the Mono Lake Excursion. The intensity values are after 20 mT and 200 °C demagnetization steps.

by the succeeding stronger fields because the directions exhibited below the marker ash are much more scattered and are not strikingly different from those of the recent field; hence they do not rule out ordinary viscous remagnetization. However, the same overprinting mechanism, but operating to a somewhat lesser extent than at the southeastern shore, can account very nicely for the absence of the expected directional swing in these records. Thus, it is reasonable to suppose as well that some sedimentary records of geomagnetic reversals may have suffered similar overprinting of their low-field portions by the higher fields at the end of polarity transitions, and also by relatively high-field episodes that may occur within transitions^{19,20}. By themselves, the three MLE records from the southeastern shore of Mono Lake appear quite acceptable, in that they meet or exceed the standards usually applied for consistency within hand samples and between sections, undisturbed sedimentary fabric, simplicity of magnetic mineralogy and quality and consistency of AF and thermal demagnetization results. Our next goal is to identify the critical properties that make the southeastern-shore sediments susceptible to such overprinting, which in turn should

help us to identify other rocks similarly incapable of carrying high-quality records of rapid field changes. □

Received 16 June; accepted 25 October 1993.

1. Rochette, P. *Earth planet. Sci. Lett.* **98**, 33–39, (1990).
2. Van Hoof, A. A. M. & Langereis, C. G. *Nature* **351**, 223–225 (1991).
3. Valet, J.-P., Tucholka, P., Courtillot, V. & Meynadier, L. *Nature* **356**, 400–407 (1992).
4. Langereis, C. G., van Hoof, A. A. M. & Rochette, P. *Nature* **358**, 226–230 (1992).
5. Weeks, R., Fuller, M., Laj, C., Mazaud, A. & Herrero-Bervera, E. *Geophys. Res. Lett.* **19**, 2007–2010 (1992).
6. Denham, C. R. & Cox, A. *Earth planet. Sci. Lett.* **13**, 181–190 (1971).
7. Liddicoat, J. C. & Coe, R. S. *J. geophys. Res.* **84**, 261–271 (1979).
8. Liddicoat, J. C. *Geophys. J. Int.* **108**, 442–452 (1992).
9. Lajoie, K. R. *thesis*, Univ. California at Berkeley (1968).
10. Lund, S. P., Liddicoat, J. C., Lajoie, K. R., Henyey, T. L. & Robinson, S. W. *Geophys. Res. Lett.* **10**, 1101–1104 (1988).
11. Negrini, R. M. & Davis, J. O. *Quat. Res.* **38**, 46–59 (1992).
12. Barbetti, M. & Flude, K. *Nature* **279**, 202–205 (1979).
13. Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. *Nature* **345**, 405–410 (1990).
14. Lovlie, R. *Terra Nova* 5, Abstr. Suppl. No. 1, **87**, (1993).
15. Liddicoat, J. C. *thesis*, Univ. California at Santa Cruz (1976).
16. Valet, J. P., Tauxe, L. & Clark, D. R. *Earth planet. Sci. Lett.* **87**, 463–472 (1988).
17. Levi, S. & Karlin, R. *Earth planet. Sci. Lett.* **92**, 219–233 (1989).
18. Negrini, R. M., Davis, J. O. & Verosub, K. L. *Geology* **12**, 643–646 (1984).
19. Shaw, J. *Geophys. J. R. astr. Soc.* **40**, 345–350 (1975).
20. Hoffman, K. A. *Nature* **359**, 783–794 (1993).

ACKNOWLEDGEMENTS. This work was supported by the US National Science Foundation, ARCO Oil and Gas and the US Geological Survey.

Evidence from borehole samples for the role of accessory minerals in lower-crustal conductivity

A. Duba*, S. Heikamp†, W. Meurer†, G. Nover† & G. Will†

* Lawrence Livermore National Laboratory, Livermore, California 94550, USA

† Mineralogisches Institut, Universität Bonn, Poppelsdorfer Schloss., Bonn, Germany

THE high electrical conductivity of the lower continental crust, as inferred from geophysical measurements¹, has defied a simple explanation. Both pore-saturating brines² and interconnected carbon films^{3–5} have been proposed as conducting pathways, but their relative importance remains uncertain, and may vary from place to place⁶. So far, attempts to address this question in the laboratory have had to rely on samples of metamorphic rock collected from surface exposures^{7–10}. Although these rocks almost certainly originated in the lower crust, chemical alteration during their transport to, and residence at, the surface is likely to have affected their conducting properties. Here we report conductivity data for pristine crustal rocks recovered from depths of up to 4.6 km in the KTB boreholes in southern Germany. Our results show that the electrical conductivity of accessory phases such as Fe–Ti oxides and sulphides can enhance, or even surpass, the high conductivity produced by saline fluids.

We have measured the electrical conductivity at hydrostatic pressures up to 250 MPa and as a function of frequency between 1 kHz and 1 MHz, of randomly selected metamorphic rocks as follows: a selection taken as cores from depths between 4,149 and 4,602 m in the *Hauptbohrung* borehole, and one sample taken at 1,858 m depth in the *Vorbohrung* borehole, both at the KTB site in Windischeschenbach, Germany. For comparison purposes, metamorphic rock (an amphibolite, number 769 in Table 1) was selected from a surface exposure near Spessart, Germany. All rocks measured were either gneisses or amphibolites with porosity of 0.3–0.8 volume %. Measurements were made at a nominal room temperature of 25 °C on saturated

rocks that had been evacuated and then back-saturated with 1 M NaCl solution following the method described in Llera *et al.*¹¹ Figure 1 shows the experimental apparatus used. Figure 2a shows the highly anisotropic conductivity measured, as a function of pressure, both parallel and at an angle to the plane of foliation on a core taken from 4,149 m in the Hauptbohrung. Electrical conductivity in the direction parallel to foliation decreases with pressure, as is commonly observed for crystalline and metamorphic rocks^{7,8}. But conductivity in the direction at an angle to the foliation increases with pressure, after decreasing with pressure in the first 10 MPa or so (where the loss of the conductive pore-saturant due to crack closure has a greater effect than the reconnection of highly conductive accessory phases that occurs at higher pressure).

Figure 2b shows the conductivity of all rocks listed in Table 1 up to 250 MPa, normalized to the value measured in the experimental assembly before pressurization (also shown in Table 1). In all the samples from the KTB site, even when pressure decreases conductivity, the effect of pressure on the conductivity of these saturated rocks is smaller than that normally observed in metamorphic rocks from surface exposures^{7,8} (see, for example, the data for sample 769 in Fig. 2b).

What is responsible for this unusual behaviour in metamorphic rocks fresh from the core barrel at the KTB site? Examination of the mineralogical and petrological data (from the KTB¹²) indicates that there can be several accessory phases present that have high conductivity, namely graphite, ilmenite,

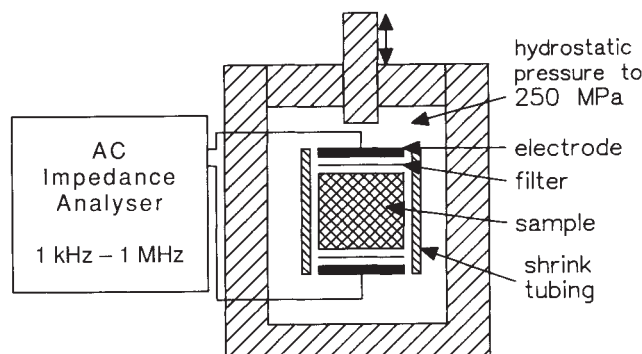


FIG. 1 Schematic diagram of experimental apparatus. The rock sample, ~30 mm in diameter and nominally 20 mm long, is sandwiched between stainless-steel electrodes and filter papers saturated with 1 M NaCl solution. The side of the electrodes in contact with the filter paper has concentric grooves, with volume more than sufficient to contain the pore saturant in the rock sample and filter paper. The assembly is held together by a shrink tube which acts also to isolate the rock from oil, the hydrostatic-pressure medium. (AC, alternating current).

pyrite and magnetite. These accessory phases and their conductivities are given in Table 1. In addition, extensive mineralization (mica, chlorite and zeolite) occurs along the foliation. Analytical electron microscopic examination of the samples clearly shows broken laths of ilmenite, as seen in the lower portion of Fig. 3, and many small inclusions of this mineral throughout the foliation direction. We propose that the effect observed here is caused by pressure acting to reconnect highly conducting films of ilmenite that had been broken by crack growth. (This crack growth occurred during the coring operation and the subsequent pressure and temperature decrease during the quick journey to the surface in the core tube.) This hypothesis is further strengthened by the observation that the frequency dispersion of conductivity decreased considerably with pressure, indicating a more metallic conductor than existed at atmospheric pressure. The conductivity of ilmenite is more than an order of magnitude higher than that of the 1 M NaCl aqueous solution used as a pore saturant. Thus, as pressure increases, the conductivity loss as the saturant is forced from the fracture porosity is more than offset by the conductivity gain as broken strands of ilmenite are reconnected. Using the data in Table 1, the cross-sectional diameter of reconnected ilmenite required to explain the increase in conductivity observed in Fig. 2b is $<10 \mu\text{m}$ —not a large amount for rocks containing as much as 1% ilmenite as accessories if the methodology of Duba and Shankland⁶ is used.

We suggest that the reason the effect of pressure on the conductivity of the samples differs from that observed in similar rocks⁷ from surface exposures is because of the presence of the highly reactive, conductive phases listed in Table 1. These phases are subject to oxidation, weathering and retrograde metamorphism during the slow journey to the surface that befalls every rock in normal denudation processes. These intergranular phases, and their freshly broken surfaces resulting from pressure and temperature decrease, are easily attacked by high-temperature saline fluids and oxygen-bearing near-surface water. The alteration products that ensue may also be highly conductive, but conductivity measurements on such rocks are usually discounted because it is easily argued that the oxidized and weathered phases do not exist at depth^{13,14}. Deep drilling projects thus offer the possibility to obtain fresh core from the interior.

We can use these data on rocks fresh from shallow crustal depths to address the controversy about the cause of the high conductivity in the crust—fluids or carbon.^{2,4,15} Because the rocks at the KTB site are almost all saturated with a 1 M NaCl-equivalent-fluid and because several fluid-producing zones have

TABLE 1 Conductivity data

Rock number	Type		σ (S m^{-1})
H001E	Amphibolite	$\triangle$ (15°)	2.24×10^{-3}
		$\blacktriangle$	4.48×10^{-3}
H003	Amphibolite	$\square$ (30°)	6.83×10^{-3}
		$\blacksquare$	4.99×10^{-3}
H005	Amphibolite	$*$ (45°)	31.66×10^{-3}
		$+$	13.08×10^{-3}
H007	Amphibolite	$\star$ (35°)	8.20×10^{-3}
		$\star$	2.04×10^{-3}
V 420	Gneiss	$\circ$	56.46×10^{-3}
		$\bullet$	17.61×10^{-3}
		$\blacklozenge$	3.77×10^{-3}
769	Amphibolite	$\oplus$	3.09×10^{-3}
(Surface exposure)		$\oplus$	0.94×10^{-3}
		$\oplus$	1.40×10^{-3}
Conducting phases			
Graphite	Accessory		10^4 – 10^7
Ilmenite	Accessory		10^2
Magnetite	Accessory		10^4
Pyrite	Accessory		10^3
Pyrrhotite	Accessory		10^5

Data for rocks from KTB cores, amphibolite 769 from a surface exposure near Spessart, and accessory phases. The conductivity value given is the value measured at 1 kHz for rocks saturated with 1 M NaCl solution in the experimental apparatus at room temperature and atmospheric pressure. The symbols used in Fig. 2b are given here for ease of comparison. The first symbol given for all Hauptbohrung rocks (identified by an H prefix) is for conductivity measured at an angle to the plane of foliation (the angle between the normal to the foliation plane and the measurement direction is given in parentheses after the symbol); the second symbol is for conductivity measured in the plane of foliation. For samples V420 and 769, the first two values given are for orthogonal measurements within the plane of foliation and the last value is perpendicular to this plane. Normalized values of conductivity for the two measurements in the foliation plane overlap exactly for sample 769 (Fig. 2). The lack of overlap for the equivalent measurements on sample V420 reflects anisotropic conductivity within the foliation plane, and is probably due to a preferred orientation of the conducting accessory phases within this plane. The data for all accessory phases are from ref. 18 except for graphite, which is from ref. 6.

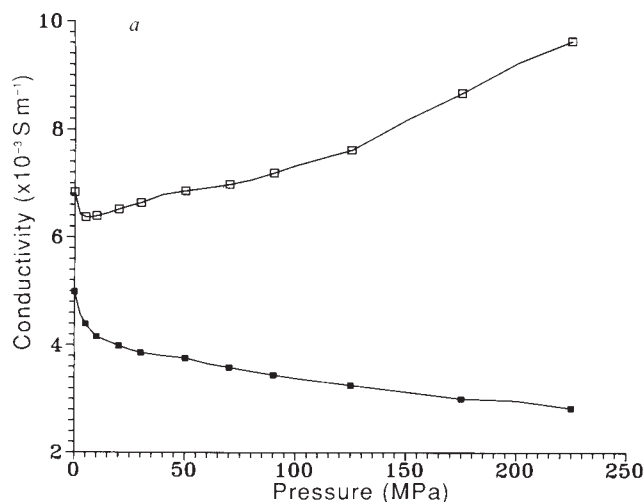
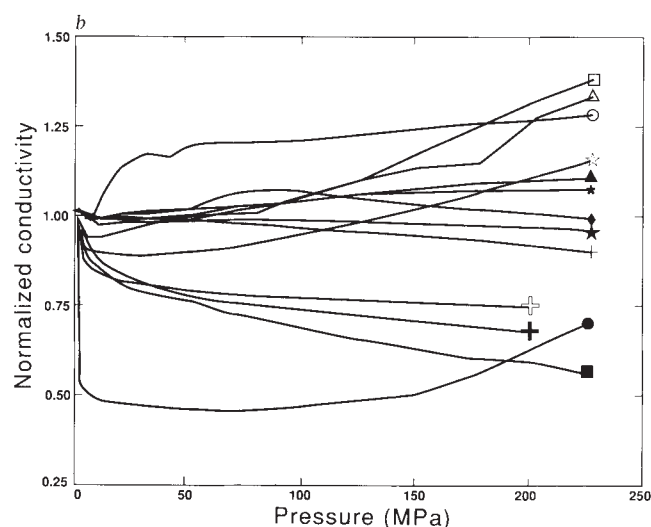


FIG. 2 *a*, Electrical conductivity of amphibolite H003 from a depth of 4,149 m in the KTB *Hauptbohrung* borehole, measured at an angle of 30° with respect to the normal of the foliation plane (□) and in the



plane of foliation (■). *b*, Normalized conductivity as a function of pressure. See Table 1 for symbols used. The density of data points is similar to *a*.

FIG. 3 Analytical electron microscope view of a portion of sample V420 showing broken laths of highly conducting ilmenite in the foliation direction of the rock. (Scale bar, 40 µm).



been observed down to ~7 km (with at least one zone between 6 and 7 km above hydrostatic pressure), we suggest that conduction in the crust in this region is the sum of highly conducting intergranular phases and ubiquitous saline fluids. By analogy, zones of high conductivity in the upper mantle could be the product of a combination of isolated good conductors joined by a partial melt of lower conductivity than the accessory phases, but of higher conductivity than most silicates at similar temperatures¹³.

There are some regions (like that around Münsterland in northern Germany) where one can say that the good conductor at depth in the crust is the result of the presence of carbon¹⁶; in other regions, the source of high conductivity is hot, highly saline geothermal brine¹⁷. We feel the results reported here demonstrate that the general case in the crust and mantle is likely to be a mixture of contributions from fluids (including melts) that connect films or isolated crystals of conductive phases such as graphite, oxides and sulphides. □

Received 21 April; accepted 30 September 1993.

1. Jones, A. G. in *The Lower Continental Crust* (eds Fountain, D. M., Arculus, R. J. & Kay, R. W.) 81–143 (Elsevier, New York, 1992).
2. Bailey, R. C., Craven, J. A., Macnae, J. C. & Polzer, B. D. *Nature* **340**, 136–138 (1989).
3. Glover, P. W. J. & Vine, F. J. *Nature* **360**, 723–726 (1992).
4. Frost, B. R., Fyfe, W. S., Tazaki, K. & Chan, T. *Nature* **340**, 134–136 (1989).
5. Losecke, W., Knödel, K. & Müller, W. *Geophys. J. R. astr. Soc.* **58**, 169–179 (1979).
6. Duba, A. G. & Shankland, T. J. *Geophys. Res. Lett.* **9**, 1271–1274 (1982).
7. Lee, C. D., Vine, F. J. & Ross, R. G. *Geophys. J. R. astr. Soc.* **72**, 353–371 (1983).
8. Brace, W. F., Orange, A. S. & Madden, T. R. *J. geophys. Res.* **70**, 5669–5678 (1965).
9. Katsube, T. J., Mareschal, M. & Aucoin, F. *Geol. Surv. Pap. Can.* **91-E**, 257–265 (1991).
10. Katsube, T. J., Scromeda, N., Mareschal, M. & Bailey, R. C. *Geol. Surv. Pap. Can.* **92-1E**, 225–236 (1992).
11. Llera, F. J., Sato, M., Nakatsuka, K. & Yokoyama, H. *Geophysics* **55**, 576–585 (1990).
12. Lich, S., Duyster, J., Godizart, G., Keyssner, S. & de Wall, H. in KTB Report **92-2** (eds Emmertmann, R., Dietrich, H. G., Lauterjung, J. & Wöhrli, Th.) B1–B83 (Schweizerbart'sche, Stuttgart, 1992).

13. Duba, A. *Acta Geodaet. Geophys. et Montanist. Acad. Sci. Hung.* **11**, 485–495 (1976).
14. Duba, A. in *High Pressure Researches in Geoscience* (ed. Schreyer, W.) 375–281 (Schweizerbart'sche, Stuttgart, 1982).
15. Shankland, T. J. *Nature* **340**, 102 (1989).
16. Duba, A., Huenges, E., Nover, G., Will, G. & Jödicke, H. *Geophys. J.* **94**, 413–419 (1988).
17. Lin, W. & Daily, W. J. *Geophys. Res.* **93**, 13047–13056 (1988).
18. Olhoeft, G. R. in *Physical Properties of Rocks and Minerals* (eds Touloukian, N. S., Judd, W. R. & Roy, R. F.) 257–328 (McGraw-Hill, New York, 1981).

ACKNOWLEDGEMENTS. We thank T. J. Shankland for critically reading the manuscript, S. Siegesmund for analytical electron microscopy and A. Kontry for information on grain-boundary phases in KTB rocks. A. D. thanks colleagues in Bonn and the KTB for their assistance, and the Alexander von Humboldt Stiftung for the opportunity to work in Germany. This work was partially supported by the US Department of Energy at the Lawrence Livermore National Laboratory.

Oxidation of humic substances by manganese oxides yields low-molecular-weight organic substrates

William G. Sunda* & David J. Kieber†

* National Marine Fisheries Service, 101 Pivers Island Road, Beaufort, North Carolina 28516, USA

† SUNY, College of Environmental Science and Forestry, Chemistry Department, Syracuse, New York 13210, USA

MANY bacteria oxidize thermodynamically unstable manganese(II) to Mn oxides and deposit the oxides on their surfaces^{1,2}, a process that appears to account for most Mn oxidation in natural waters³⁻⁵ and sediments⁶. Among the reasons that have been proposed for the evolutionary selection of this process are protection from damage by toxic metals and oxygen species, protection from ultraviolet light, and strengthening of the bacterial sheath or capsule^{1,7}. Mn oxides may promote harmful free radical reactions, however, and marine Mn-oxidizing bacteria are particularly susceptible to photoinhibition⁸. Here we report that Mn oxides lyse complex humic substances, which in general cannot be used by microorganisms directly⁹⁻¹¹, to form low-molecular-weight organic compounds that can be used as substrates for microbial growth. Mn-oxidizing bacteria may thus be able to use the carbon pool in humic substances, which represent one of the largest organic reservoirs in natural waters, sediments and soils.

Mn oxides are reductively dissolved by humic and fulvic acids^{12,13} and by certain simple organic compounds such as phenols and organic acids (such as pyruvate and oxalate)^{14,15}. The products from the oxidation of humic substances by Mn oxides have not been investigated, although it is likely that a variety of low-molecular-weight (LMW) carbon compounds will be produced (in addition to CO₂ and CO) because of the heterogeneous nature of these substances. To test this hypothesis, we added synthesized Mn oxides to filtered estuarine water containing a high concentration (4.1 mg l⁻¹) of natural organic carbon and to low organic sea water from the Gulf of Mexico amended with different additions of isolated fulvic acids. Following addition of the oxides, the suspensions were filtered through 0.2 µm-pore nylon filters and the concentrations of LMW aldehydes and ketones¹⁶ and α-ketoacids¹⁷ were measured in the filtrates by high-performance liquid chromatography (HPLC).

A variety of LMW carbonyls were produced from the oxidation of dissolved natural organic matter (NOM) by the Mn oxides, including pyruvate, acetone, formaldehyde, acetaldehyde (in descending order of importance), and an unknown (Fig. 1). In experiments with fulvic acids (Fig. 2), the initial rates of production of pyruvate and acetaldehyde conformed to a Langmuir adsorption isotherm:

$$dP/dt = \frac{V_{\max}[FA]K}{[FA]K + 1} \quad (1)$$

where dP/dt is the formation rate of product, $[FA]$ is the concentration of fulvic acids, K is a pseudoequilibrium adsorption constant, and $V_{\max}$ is the maximum reaction rate achieved when all adsorption sites are saturated. A similar hyperbolic rate law has been found to describe the rate of reduction of Mn oxides by fulvic acids¹³. The fit of the reaction rate to an adsorption isotherm supports the general model of rapid adsorption of organic molecules onto Mn oxides followed by slower electron transfer at the oxide surface to produce Mn(II) and organic oxidation products¹⁸.

Nonlinear least-square fits of our fulvic acid rate data to equation (1) yielded $V_{\max}$ values of 36 ± 2 (±s.d.) and

2.4 ± 0.1 nM h⁻¹ and values for K of 0.18 ± 0.03 and 0.13 ± 0.02 l mg⁻¹ (at pH 8.0) for pyruvate and acetaldehyde, respectively (Fig. 2). These K values are in the same range as those (0.16 at pH 4 and 0.023 l mg⁻¹ at pH 7.1) determined in ref. 13 for the rate of reduction of Mn oxides by the fulvic acids in 0.05 M NaCl. Comparisons between our results and those from ref. 13 are particularly appropriate because both studies used oxides synthesized by similar methods, and both used samples of the same Suwanee River fulvic acid preparation. A similarity in K values would be expected because both Mn(II) and the two LMW carbonyls are produced from the same reaction: oxidation of adsorbed fulvics by the Mn oxides. Our slightly higher K values (at pH 7-8) may reflect increased fulvic acid adsorption onto Mn oxides with increased salinity as a result of electrostatic screening of negative charge.

A time-course experiment in filtered water from the mouth of the Shark River estuary, Florida, showed a rapid production of formaldehyde and acetaldehyde during the first few hours after oxide addition followed by decreasing rates thereafter and little production after 21 h (Fig. 3). In a second experiment in this water, the net soluble yields of these two compounds after 14 h increased with oxide concentration, and were linearly related ($r^2 = 0.995$ and 0.993) to the amount of NOM removed due to reaction with the oxides as quantified by the decrease in NOM absorbance in sample filtrates at 300 nm (Fig. 4). Yields for pyruvate and acetone showed considerably different behaviour; their net production reached maxima and then decreased with increasing oxide concentrations (Fig. 4). These data are consist-

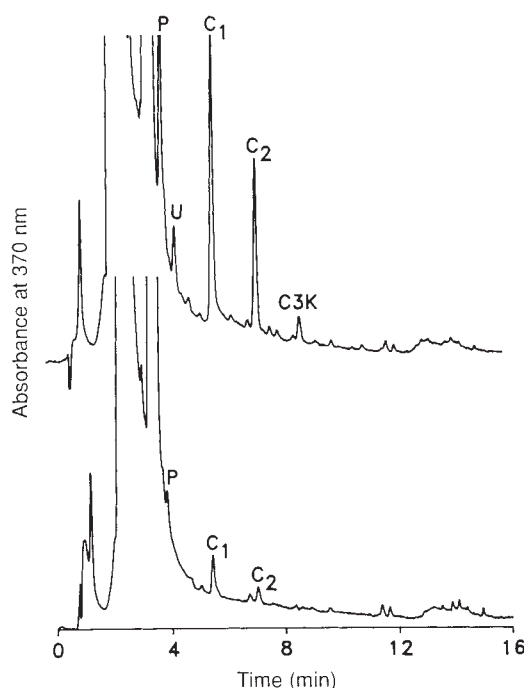


FIG. 1 HPLC chromatograms for dissolved LMW carbonyls in estuarine water initially (lower) and after 14.3 h exposure to 1 mM Mn oxides at pH 7.7 (upper). The peaks indicated are for the hydrazone derivatives¹⁷ of pyruvate (P), formaldehyde (C₁), acetaldehyde (C₂), acetone (C₃K), and one unknown (U). Chromatograms used to quantify pyruvate (using a mobile phase of higher pH¹⁹) are not shown. The 22‰ salinity water was collected from the estuary of the Shark River, which drains in part the Everglades Swamp in southwestern Florida. It was stored in a high density polyethylene carboy in the dark at 4 °C for 10 months and was filtered through a 0.2 µm pore Nylon filter just before use. Manganese oxides were prepared from alkaline permanganate oxidation of MnCl₂ (ref. 21), a procedure that yields oxides of Mn oxidation state 3.8-4.0 (refs 13, 21). All experiments were done at 23 ± 1 °C.

ent with subsequent oxidation of pyruvate¹⁵ and actone on the oxide surface following their production. The linear yield behaviour of formaldehyde and acetaldehyde may be related to their greater resistance toward oxidation and/or to a greater tendency of these C₁ and C₂ compounds to desorb from the oxide following their formation.

Results of the above experiment were used to estimate yields for carbonyl products as a percentage of natural organic carbon (NOC). Yields were computed from initial slopes of relationships between molar concentration of product and delta absorbance (Fig. 4) multiplied by the number of carbon atoms per compound. The resulting values were divided by the ratio of NOC to absorbance for the estuarine water (2.2 mmol C l⁻¹/absorbance at 300 nm). In these computations we assumed that the NOC adsorbed and/or reacted is proportional to the absorbance lost from solution, a reasonable assumption given the linear relationship between concentrations of C₁ and C₂ products and delta absorbance. We estimate that 0.32, 0.128, 0.0079 and 0.0067% of the organic carbon reacted was converted to pyruv-

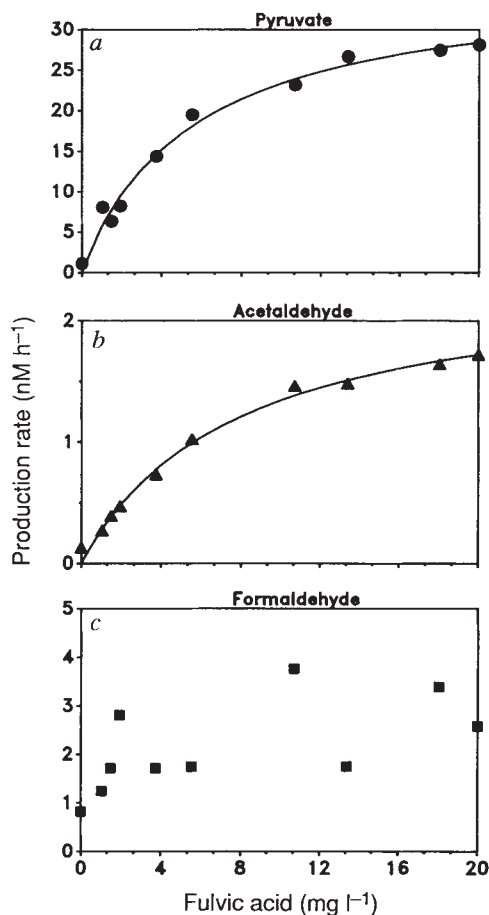


FIG. 2 Initial formation rates of a, pyruvate, b, acetaldehyde, and c, formaldehyde at pH 8.0 ± 0.1 in Gulf of Mexico water containing 1 mM oxides and 0–20 mg l⁻¹ fulvic acids isolated from the Suwannee River, Georgia¹⁹. Rates were determined by measuring filterable carbonyl concentrations immediately after oxide addition and after 1.5 h reaction. They are plotted as a function of the fulvic acid concentrations in sample filtrates at 1.5 h, as measured spectrophotometrically at 300 nm. Data for pyruvate and acetaldehyde fit well ($r^2 = 0.996$ and 0.997 , respectively) to model curves computed from equation (1) using V_{max} and K values listed in the text. Because of scatter (presumably due to random contamination), formaldehyde rate data were not modelled. Acetone formation rates could not be measured because of insufficient analytical sensitivity.

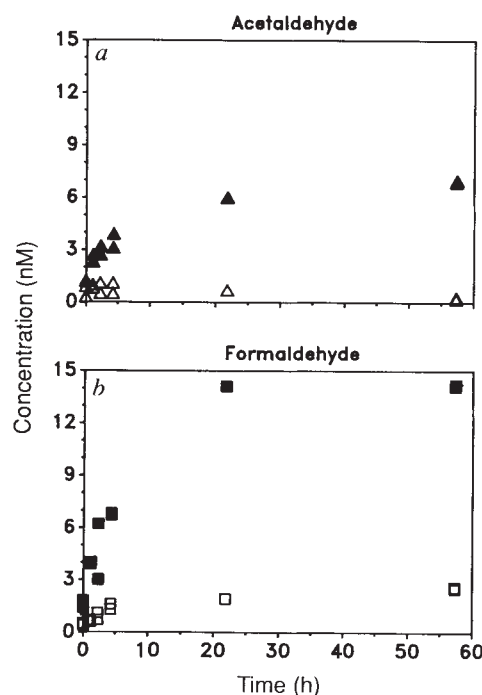


FIG. 3 Time course for formation of dissolved formaldehyde and acetaldehyde at pH 7.3 in Shark River estuarine water with (closed symbols) and without (open symbols) 1 mM added Mn oxides. α -Ketoacids were not measured.

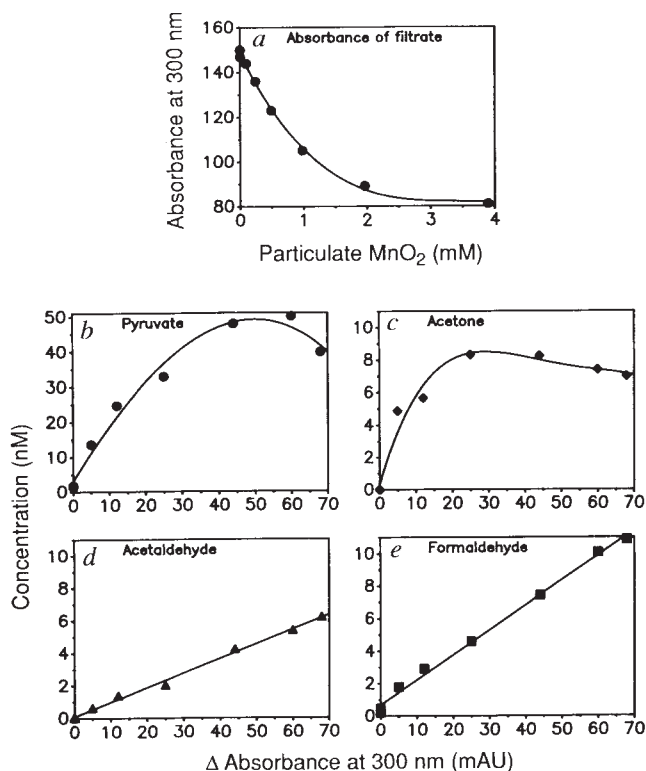


FIG. 4 Effect of added Mn oxides on loss of NOM absorbance (a) and net production of pyruvate (b), acetone (c), acetaldehyde (d) and formaldehyde (e) after 14.3 h reaction in Shark River estuarine water at pH 7.7 ± 0.1 . Carbonyl values were corrected for blank values measured immediately after oxide addition and are plotted as functions of the loss of NOM absorbance at 300 nm for each Mn oxide addition (see a). Linear regression lines are drawn for formaldehyde and acetaldehyde data.

ate, acetone, acetaldehyde, and formaldehyde, respectively. The estimated yield for all four products was 0.46%. This value probably underestimates the overall yield for these compounds because some organic matter that adsorbs to the oxides may not react, and some stable products may not desorb following formation. Also, our calculations almost certainly underestimate the overall formation of LMW organic products, because our analytical methods would not detect several major classes of likely products such as organic acids (other than α -ketoacids) and alcohols.

The dissolved NOC in most oxidizing environments consists primarily of biologically refractory compounds such as humic and fulvic acids. This material is composed of complex mixtures of moderately high-molecular-weight organics whose structural diversity defies both specific molecular identification and coherent enzymatic attack. Thus, the oxidative lysis of these compounds by Mn oxides represents a mechanism by which biologically refractory organic matter is converted into a suite of low-molecular-weight compounds that can be used as microbial substrates. Pyruvate, the major LMW product that we identified, is a major metabolic intermediary in the cell's TCA cycle, and as such, provides a ready source of both energy and carbon for growth. It is also a major product of the photolysis of humic compounds by ultraviolet light, and its photochemical production is closely coupled to its microbial use in sea water¹⁹.

In using LMW products, such as pyruvate, formed from photolysis of NOM¹⁹, microorganisms can be viewed as exploiting an external photochemical process that is otherwise beyond their control. But this is not true in the case of microbial use of the same or similar compounds formed from Mn oxide lysis of NOM because Mn oxides in most natural environments are formed by microbial catalysis³⁻⁶ and are found associated with the extracellular polymer sheath (glycocalyx) surrounding the

bacterial cells^{1,2}. Although several hypotheses have been proposed¹, the primary reason (or reasons) why the bacteria oxidize Mn (II) and deposit Mn oxides on their surfaces remains uncertain. We speculate that it may represent a means by which they can use the carbon contained in the large but biologically refractory pools of NOM, such as humic and fulvic acids. By mediating the production of Mn oxides these microbes also could play an important role in the biogeochemical cycling of organic carbon, particularly in surficial sediments where Mn oxide concentrations occur at much greater levels than those used in our experiments^{6,10}. □

Received 2 June; accepted 17 September 1993.

1. Ghiorse, W. C. *A. Rev. Microbiol.* **38**, 515-550 (1984).
2. Cowen, J. P. & Bruland, K. W. *Deep-Sea Res.* **32**, 253-272 (1985).
3. Ghiorse, W. C. in *Current Perspectives in Microbial Ecology* (eds Klug, M. J. & Reddy, C. A.) 615-621 (ASM, Washington DC, 1984).
4. Tebo, B. M. & Emerson, S. *Appl. Environ. Microbiol.* **50**, 1268-1273 (1985).
5. Sunda, W. G. & Huntsman, S. A. *Limnol. Oceanogr.* **32**, 552-564 (1987).
6. Kepkay, P. E. *Limnol. Oceanogr.* **30**, 713-726 (1985).
7. Emerson, D. & Ghiorse, W. C. *Appl. Environ. Microbiol.* **58**, 4001-4010 (1992).
8. Sunda, W. G. & Huntsman, S. A. *Limnol. Oceanogr.* **35**, 325-338 (1990).
9. Mopper, K. & Degens, E. T. in *The Global Carbon Cycle* (eds Bolin, B., Degens, E. T., Kemp, S. & Kettner, P. (Wiley, New York, 1979).
10. Geller, A. *Limnol. Oceanogr.* **31**, 755-764 (1986).
11. Amador, J. A., Alexander, M. & Zika, R. G. *Appl. Environ. Microbiol.* **55**, 2843-2849 (1989).
12. Sunda, W. G., Huntsman, S. A. & Harvey, G. R. *Nature* **301**, 234-236 (1983).
13. Waite, T. D., Wrigley, I. C. & Szymczak, R. *Environ. Sci. Technol.* **22**, 778-785 (1988).
14. Stone, A. T. & Morgan, J. J. *Environ. Sci. Technol.* **18**, 617-624 (1984).
15. Stone, A. T. *Geochim. cosmochim. acta* **51**, 919-925 (1987).
16. Kieber, R. J. & Mopper, K. *Environ. Sci. Technol.* **24**, 1477-1481 (1990).
17. Kieber, D. J. & Mopper, K. *Mar. Chem.* **21**, 135-149 (1987).
18. Stone, A. T. & Morgan, J. J. *Environ. Sci. Technol.* **18**, 450-456 (1984).
19. Kieber, D. J., McDaniel, J. & Mopper, K. *Nature* **341**, 637-639 (1989).
20. Aller, R. C. *Phil. Trans. R. Soc.* **331**, 51-68 (1990).
21. van den Berg, C. J. M. & Kramer, J. R. *Analyst. Chem. Acta* **106**, 113-120 (1979).

ACKNOWLEDGEMENTS. We thank A. Sharp for synthesis of Mn oxides and B. Yocis for help in conducting the experiments. This work was supported by the Oceanic Chemistry Program of the Office of Naval Research.

A genetic analysis of senescence in *Drosophila*

Kimberly A. Hughes*† & Brian Charlesworth‡

* Committee on Evolutionary Biology, and ‡ Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois 60637, USA

Two attractive theories for the evolution of senescence are based on the principle that the force of natural selection decreases with age¹⁻⁵. The theories differ in the type of age-specific gene action that they assume. Antagonistic pleiotropy²⁻⁵ postulates that pleiotropic genes with positive effects early in life and negative effects of comparable magnitude late in life are favoured by selection, whereas genes with the reverse pattern of action are selected against. Mutation accumulation^{1,3-5} assumes that deleterious mutant alleles with age-specific effects will equilibrate at a lower frequency if their effects are expressed early rather than late in life. Explicit models demonstrate that both mechanisms can lead to the evolution of senescent life histories under reasonable conditions³⁻⁵. Antagonistic pleiotropy has gained considerable empirical support⁴⁻⁶, but the evidence in support of mutation accumulation is more sparse^{4,5,7}. Here we report that the genetic variability of mortality in male *Drosophila melanogaster* increases greatly at very late ages, as predicted by the mutation accumulation hypothesis³⁻⁵. The rate of increase in mortality with age exhibits substantial genetic and environmental variability. This result provides a possible explanation for recent observations of non-increasing mortality rates in very old flies^{8,9}.

Forty wild-type third chromosomes were extracted from a large, outbred laboratory population of *D. melanogaster* (the IV

stock¹⁰). Wild-type male progeny produced by intercrossing the different extraction lines were assayed for age-specific survivorship (Table 1 and Fig. 1) to test a prediction of the mutation-accumulation theory of senescence^{1,3-5}. If deleterious mutant alleles have effects confined to specific ages, the equilibrium gene frequency for deleterious mutations should increase with the age of gene effect under a wide range of genetic and demographic conditions³⁻⁵. The mean value of fitness-related traits will therefore decrease with age and their genetic variability will increase. The decline in the mean of such a trait will contribute to senescence (defined here as a decline in fitness components with age). Both the additive genetic component of variance (V_A) and dominance component (V_D) are expected to increase with age under the mutation accumulation hypothesis (because for traits affected by rare deleterious alleles, both components increase with increasing mutant allele frequency^{11,12}).

Analyses of mortality over 3-week periods indicate that the genetic (V_A and V_D) and environmental (V_E) components of variance for this character increase dramatically with age (Table 1). All genetic components are statistically indistinguishable from zero during the early and intermediate periods, but are significantly greater than zero for the late time interval. It is clear from Table 1 that, at weeks 9 to 11, all five estimates of V_A , V_D and V_E from independent blocks of the experiment are much larger than corresponding estimates for earlier intervals. This result is not due to an inability to detect variation for low mortality rates, because the variance components for mortality during weeks 1 to 9 are very small compared to those for weeks 9 to 11, whereas the mean mortality rates over these two periods are comparable (for weeks 1 to 9, mean mortality rate $\mu = 14.6\%$, and the standardized variance components are $V_A = 3.94 \times 10^{-4}$, $V_D = 9.34 \times 10^{-4}$, $V_E = 14.7 \times 10^{-4}$). A large accumulation of mutational effects at very late ages, as exhibited by this population of flies, is predicted by theory¹³.

† Present address: Chicago Zoological Society, Brookfield, Illinois 60513, USA.

Although there is ample evidence for genetic variation in lifespan and age-specific fertility^{4,5,14}, there appear to be no data on the extent of genetic variation in the rate of senescence itself. We used the Gompertz model of mortality to examine this question with our data. This model derives from the empirical tendency for mortality rates to increase exponentially with age in many animal species^{14,15}. Gompertz parameters are estimated as the slope and intercept of the regression line relating log mortality to age. The intercept of this line represents a baseline, or 'intrinsic' mortality rate for organisms that are not subject to the effects of senescence. The slope of the line estimates the rate of increase in mortality with age, and provides a measure of the rate of senescence. That is, the slope estimates the accelerating effect of senescence on the risk of dying in a given time interval¹⁴. Because the genotypes used in this experiment exhibited very low mortality before 6-weeks post-eclosion, and mortality generally increased monotonically after this period, the Gompertz parameters were estimated from the mortality rates observed after the first 42 days of adult life.

The additive genetic variance for the slope of the Gompertz curve was significantly different from zero, but that for the Gompertz intercept was not (see Fig. 1 legend). This suggests that

much of the heritable variability for lifespan in this population is due to variability in rates of senescence, and not to variability in baseline mortality rates. V_D estimates for the slope and intercept were not significant. Estimates of the coefficient of environmental variation were large and highly significant for both Gompertz parameters (0.210, $P < 0.001$, for the slope, and 0.270, $P = 0.03$ for the intercept). An increase in environmental variation with age has also been observed for female fecundity¹⁶.

These results represent the first demonstration to our knowledge of heritable variation in rates of senescence within a population of organisms. Even if the Gompertz parameterization is not a completely general description of patterns of mortality^{8,9}, the results strongly suggest that substantial heritable genetic variability exists for the change in mortality rates with age. The large amounts of both genetic and non-genetic variation in the rate of senescence provide a possible explanation for the results of two recent studies of mortality rates in very large cohorts of med flies and fruit flies^{8,9}. Both these studies suggest that cohort mortality rates cease to accelerate near the end of the lifespan of the oldest surviving flies. Individual variation in the rate of increase of mortality with age, whether genetic or environmental, would lead to the appearance of lower rates of increase at

FIG. 1 Regression lines of log (mortality + 1) on days since eclosion for 5 blocks, each consisting of 16 different sets of heterozygous third-chromosome genotypes. A third-chromosome balanced lethal system was constructed by placing *TM6/Sb* on an *IV* genetic background, producing a stock that is $+/+$, $+/+$, *TM6/Sb* for the X, second and third chromosomes, respectively. The *TM6* chromosome functions as an effective suppressor of crossing-over for the third chromosome²² and *Sb* is lethal in homozygous condition. The balancer stock is of the P/R cytotype, and should not induce hybrid dysgenesis²³ when crossed to flies from the *IV* stock. This balanced-lethal system was used to produce lines of flies carrying identical third chromosomes, heterozygous over *TM6*, as in ref. 20. Extraction lines were subdivided and maintained as two independent sublines. In any given experimental cross, flies from only one subline were used; the other subline was used in the replicate cross for that genotype. Common environmental effects and effects of common genetic background within a subline contribute to the variance within lines, not to the variance among lines. The 'environmental' variance may therefore be somewhat inflated. The effect is probably small, because many females from the balancer stock were used as parents in the first 2 generations of the extraction procedure (the variance due to the genetic background is $1/n$ times the corresponding genetic variance, where n is the number of maternally derived diploid genotypes per line). The 40 extraction lines were intercrossed in a North Carolina II crossing scheme²⁵, using 5 independent blocks of 4 maternal lines and 4 paternal lines (and reciprocal crosses) to produce 16 different heterozygous third-chromosome genotypes within each block. The additive genetic coefficient of variation (CV_A , equal to the ratio of $\sqrt{V_A}$ to the mean²⁴) for the slopes, pooled over 5 blocks, is 0.203, and is significantly greater than 0 ($P < 0.01$, $t = 4.12$). CV_A for the intercepts is 0.091 ($P = 0.16$, $t = 1.15$). Significance tests are based on 1-tailed t -tests of the variance components in 5 independent blocks.

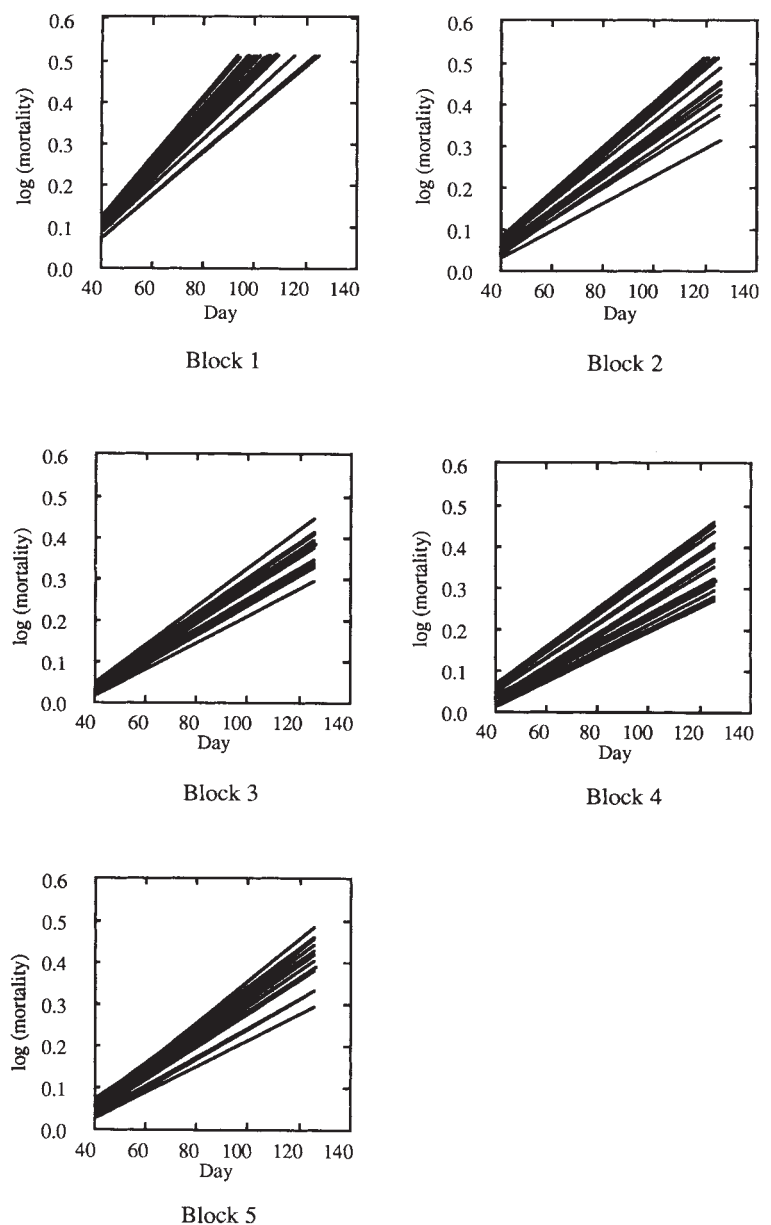


TABLE 1 Means and variance components for 3-week mortality rates at 3 ages

Block	Ecdysis to 3 weeks post-eclosion				5 to 7 weeks post-eclosion				9 to 11 weeks post-eclosion			
	μ (%)	$V_A \times 10^4$	$V_D \times 10^4$	$V_E \times 10^4$	μ (%)	$V_A \times 10^4$	$V_D \times 10^4$	$V_E \times 10^4$	μ (%)	$V_A \times 10^4$	$V_D \times 10^4$	$V_E \times 10^4$
1	0.52	0.7	-0.7	2.4	7.16	2.1	-0.2	15.6	81.5	91.4	40.2	489.6
		0.7	-0.6	2.1		2.0	-0.4	13.3		33.1	20.9	129.0
2	0.08	0.0	0.0	0.4	1.81	-0.1	1.1	9.7	19.2	60.7	37.8	64.7
		0.0	0.0	0.4		-0.1	1.3	8.2		44.0	27.3	41.7
3	0.47	-0.1	0.8	2.7	2.08	-0.5	2.4	9.6	25.1	53.8	35.9	69.1
		-0.1	0.8	2.5		-0.4	1.9	9.1		34.3	23.2	43.4
4	0.49	0.2	-0.4	2.5	0.92	0.3	-0.5	4.2	13.6	61.0	8.9	46.2
		0.1	-0.3	2.1		0.3	-0.4	3.9		47.1	7.6	34.0
5	0.48	0.0	-0.2	2.7	4.64	-0.5	2.5	21.0	27.8	34.5	31.4	251.1
		0.0	-0.2	2.4		-0.7	2.8	15.4		17.8	29.2	113.2
Pooled	0.41	0.1	0.1	1.3	3.32	0.2	1.7	7.1	33.4	59.4	30.5	89.7
		0.2	0.2	1.1		0.2	1.6	5.9		35.7	26.3	37.0
t-value		1.06	-0.20	4.89		0.73	1.65	4.96		6.88	5.70	3.59
Probability		n.s.	n.s.	<0.01		n.s.	n.s.	<0.01		<0.01	<0.01	<0.05

Blocks 1 to 5 represent partial-diallel crosses involving five independent sets of third-chromosome isogenic lines. μ is mean mortality rate, over a three-week period, of groups of flies representing many different heterozygous genotypes (see Fig. 1 legend). The mean is taken over all replicates in a block. Additive-genetic, dominance and environmental variance components for each block are given by V_A , V_D and V_E , respectively. These variance components were estimated by the synthesis method¹⁸, which produces unbiased estimates for random models¹⁹. No reciprocal effects were found, so reciprocal crosses were treated as simple replicates²⁰. Pooled estimates of variance components over all 5 blocks were calculated by the formula given in ref. 20. Within each block, the values for V_A , V_D and V_E reported on the upper line were calculated from unstandardized mortality rates; values appearing on the lower line were calculated from mortality rates that were first standardized to a block mean of 1. One-sample *t*-tests of the estimates from the 5 independent blocks indicate that, after correction for multiple tests, V_E is significantly greater than 0 at all ages, whereas V_A and V_D are significantly greater than 0 only for the latest time interval. Techniques used for estimating age-specific mortality are similar to those used in ref. 21. Males of each genotype were kept (without females) at a density of 20 per vial and were transferred weekly. Survivors were counted at each transfer until all flies were dead. To maintain constant density, dead flies were replaced by flies carrying a visible mutation. The proportion of flies dying from one transfer to the next gives an estimate of age-specific mortality for the genotype. These estimates were used to calculate mortality in 3 non-overlapping time intervals (all replicates contained some live flies throughout the experiment), and to compute Gompertz mortality parameters¹⁵ for each genotype.

advanced ages, because only individuals with low rates would survive to the end of the experiment. The large experiment in ref. 9 used highly inbred flies, which are known to be particularly sensitive to environmental effects^{11,17}. □

Received 23 July; accepted 19 October 1993.

- Medawar, P. B. *An Unsolved Problem of Biology* (Lewis, London, 1952).
- Williams, G. C. *Evolution* **11**, 398-411 (1957).
- Charlesworth, B. *Evolution in Age-structured Populations* (Cambridge Univ. Press, Cambridge, 1980).
- Rose, M. R. *The Evolutionary Biology of Aging* (Oxford Univ. Press, Oxford, 1991).
- Partridge, L. & Barton, N. H. *Nature* **362**, 305-311 (1993).
- Service, P. M. *Evolution* **47**, 387-399 (1993).
- Roper, C., Pignatelli, P. & Partridge, L. *Evolution* **47**, 445-455 (1993).
- Carey, J. R., Liedo, P., Orzco, D. & Vaupel, J. W. *Science* **258**, 457-460 (1992).
- Curtis, J. W., Fukui, H. H., Townsend, D. R. & Vaupel, J. W. *Science* **258**, 461-463 (1992).
- Charlesworth, B. & Charlesworth, D. *Heredity* **54**, 71-84 (1985).
- Falconer, D. S. *Introduction to Quantitative Genetics* (Longman, London, 1989).
- Charlesworth, B. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Andersson, M. B.) 21-40 (Wiley, Chichester, 1987).
- Charlesworth, B. *Evolution* **44**, 520-538 (1990).
- Finch, C. E. *Longevity, Senescence, and the Genome* (Univ. Chicago Press, Chicago, 1990).
- Finch, C. E., Pike, M. C. & Whitten, M. *Science* **249**, 902-905 (1990).
- Rose, M. R. & Charlesworth, B. *Genetics* **97**, 173-186 (1981).
- Wright, S. *Evolution and the Genetics of Populations Vol. 3, Experimental Results and Evolutionary Deductions* (Univ. Chicago Press, Chicago, 1977).
- Hartley, H. O. *Biometrics* **23**, 105-114 (1967).
- Searle, S. R. *Biometrics* **27**, 1-76 (1971).
- Mukai, T., Cardellino, R. A., Watanabe, T. K. & Crow, J. F. *Genetics* **78**, 1195-1208 (1974).
- Rose, M. R. *Evolution* **38**, 516-526 (1984).
- Lindsley, D. L. & Zimm, G. G. *The Genome of Drosophila melanogaster* (Academic, San Diego, 1992).
- Engels, W. R. in *Mobile DNA* (eds Berg, D. E. & Howe, M. M.) (Am. Soc. Microbiol., Washington DC, 1989).
- Comstock, R. E. & Robinson, H. F. in *Heterosis* (eds Gowen, J. W.) (Iowa State College Press, Ames, 1952).
- Houle, D. *Genetics* **130**, 195-204 (1992).

ACKNOWLEDGEMENTS. We thank S. Assimacopoulos, K. A. Dixon and K. Donohue for their help with the experiments and D. Houle, D. Hartl, M. Rose and L. Partridge for their comments on the manuscript. This work was supported by an NIH research grant to B.C. and by an NIH Genetics Training Grant to the University of Chicago.

Quantitative stereoscopic depth without binocular correspondence

Lei Liu, Scott B. Stevenson & Clifton M. Schor

School of Optometry, University of California at Berkeley, Berkeley, California 94720, USA

WHAT features in a stereogram define the disparities that lead to stereoscopic depth? The usual answer is that luminance-defined edges from the two eyes are matched and produce depth perception^{1,2}. But parts of an object may be occluded by other objects and absent from one eye's view. It was suggested that unpaired monocular elements might signal occlusion in depth, and the qualitative perception of depth associated with unmatched elements has been shown to be consistent with the geometry of occlusion³⁻⁵. We designed a stereogram that simulates a particular occlusion situation: an opaque white rectangle is stereoscopically

in front of a large black rectangle pasted on a white background. The position of the occluder is adjusted so that its left edge obscures the left-hand edge of the black rectangle in the right eye view and its right edge obscures the right-hand edge of the black rectangle in the left eye view. We report here that quantitative stereopsis can be seen from this stereogram, even though there are no binocular corresponding luminance edges to match.

The basic element of the stereogram is a square bracket pattern (Fig. 1a, b). When binocularly cross-fused, a white rectangle is seen floating in front of the image plane (Fig. 1a) or lying behind the image plane (Fig. 1b). In the fused image, the outside edges of the vertical bars of one eye's bracket are aligned with the short vertical edges above and below the openings of the other eye's bracket. Notice that neither the left nor the right edge of the middle parts of the vertical bars have corresponding luminance edges in the other eye's image.

Stereopsis without point to point correspondence has been reported previously. Monocular figures defined by either subjective contours or texture borders could be combined binocularly and produce illusory surfaces in depth as if they were defined by normal luminance contours⁶⁻⁹. But the phenomenon illus-

trated in Fig. 1*a, b* cannot be explained by binocular combination of two monocularly completed white rectangles. Visual completion of the monocular pattern is not obvious. No illusory contours or brightness induction at the openings are perceived and it is difficult to perceive a single square bracket as a black rectangle occluded by a white rectangle. Only after fusing two brackets is occlusion readily perceived. Figure 1*c* shows only the upper half of Fig. 1*a*. The 'L'-shaped monocular features are even harder to complete, yet vivid depth is seen when they are binocularly fused. Even if monocular completion does occur at an unconscious level, then the rules of completion would predict that the completed portion in Fig. 1*d* should be part of the ellipse¹⁰. When this curve matches with the straight edge from the other eye, the binocular percept should be a cylindrical occluder, similar to that seen in Fig. 1*e*. But the perceived occluder in Fig. 1*d* is flat rather than cylindrical, indicating that monocular completion may not be a proper explanation.

Figure 2*a* contains a series of bracket patterns. They differ only in the widths of the vertical bars. When cross-fused, one sees a staircase of white rectangles with increasing depth from

the bottom to the top. Using test patterns shown in Fig. 2*b* inset, we quantified the perceived depth from a stereogram without binocular correspondence (the upper part) by comparing it with the perceived depth produced by a normal stereogram (the lower part). As shown in Fig. 2*b*, to match the apparent depths on the upper part, the lower part needs a binocular disparity equal to the width of the vertical bars of the square bracket.

Figure 3 illustrates a real-world situation which can produce retinal images like those in Fig. 1*a*. Note that the situation portrayed in Fig. 3 is actually ambiguous. Any white occluder that fits in the light grey area of Fig. 3*B* produces the same retinal images. In the case of object *bb*, it occludes the right edge of object *pp* and part of the white background from the left eye's view but allows the same portion of the right side of *pp* to reach the left eye as does occluder *aa*. But human observers generally perceive the particular, unique solution represented by *aa*. As shown in Fig. 2*b*, the perceived occluder is positioned at the minimum depth at which the right edges of the occluder and *pp* are aligned on the same visual direction of the left eye, and the left edges of the occluder and *pp* are aligned on the same visual

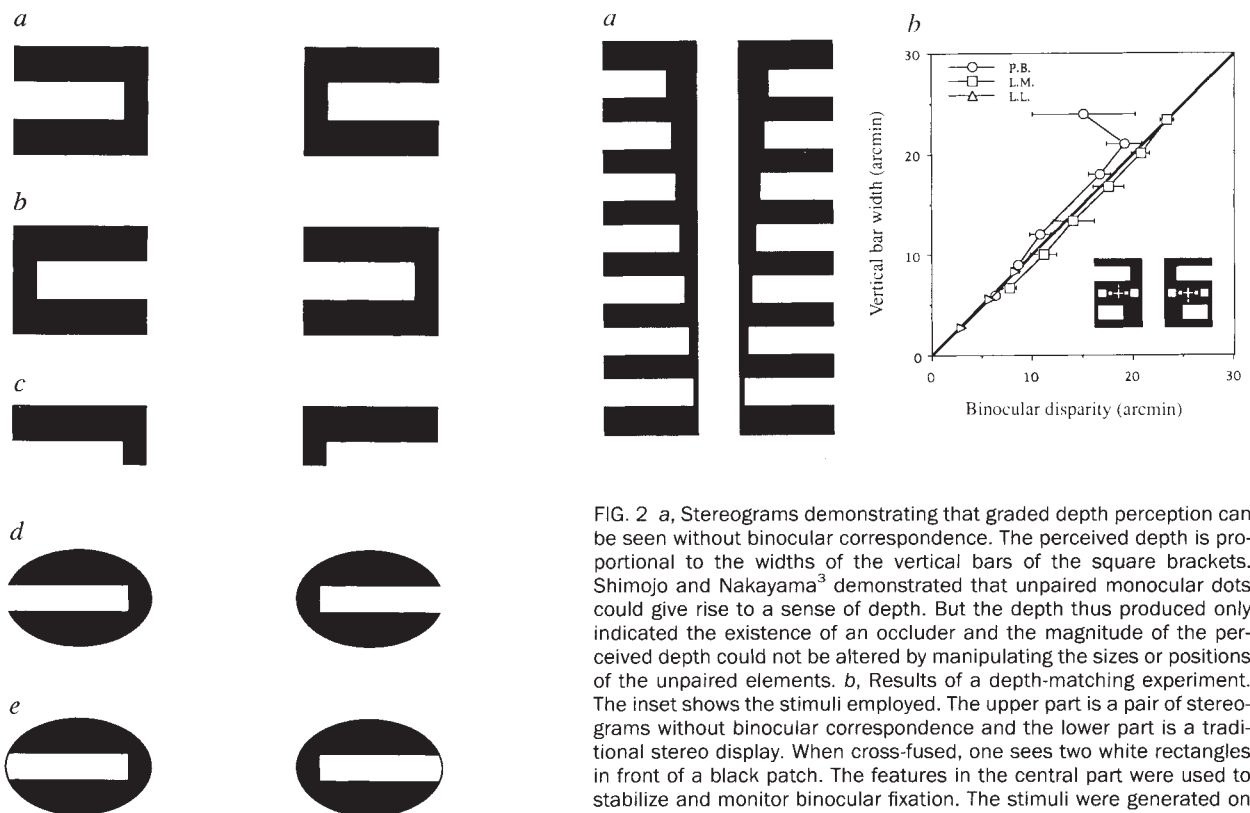
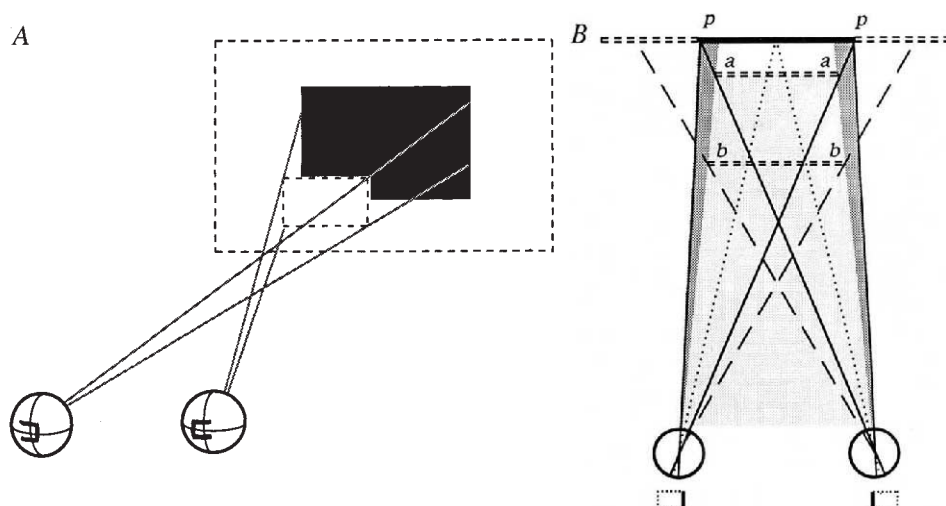


FIG. 2 *a*, Stereograms demonstrating that graded depth perception can be seen without binocular correspondence. The perceived depth is proportional to the widths of the vertical bars of the square brackets. Shimojo and Nakayama³ demonstrated that unpaired monocular dots could give rise to a sense of depth. But the depth thus produced only indicated the existence of an occluder and the magnitude of the perceived depth could not be altered by manipulating the sizes or positions of the unpaired elements. *b*, Results of a depth-matching experiment. The inset shows the stimuli employed. The upper part is a pair of stereograms without binocular correspondence and the lower part is a traditional stereo display. When cross-fused, one sees two white rectangles in front of a black patch. The features in the central part were used to stabilize and monitor binocular fixation. The stimuli were generated on the screen of an Apple Macintosh computer. For each vertical bar width in the upper part, the observer used the keyboard to adjust the binocular disparity in the lower part of the display until the two white rectangles appear to have the same depth. Three observers participated in the experiment. The viewing distances were 350, 162 and 145 cm for L.L., P.B. and L.M. Visual acuities were corrected to normal for the viewing distances. The vertical and horizontal axes represent the width of the vertical bar in the upper part and the binocular disparity in the lower part of the stimulus (defined as the difference between the width of the vertical bar on the right of the white rectangle and that on the left), respectively. Along the heavy diagonal line, binocular disparity equals the width of the vertical bar. Each data point was the average of 15 adjustments. L.M. (square symbols) could not make depth judgments when bar width was 27 arcmin because, if she kept her fixation at the centre of the display, the white areas above and below appeared double. P.B. (open circles) could make depth matches when the bar width was 24 arcmin, but only with great effort. Her results were noisy and the average perceived depth was smaller than those observed with narrower vertical bars.

FIG. 1 *a, b*, Stereograms without binocular correspondence. When cross-fused, a white rectangle is seen in front of a black patch in *a* and behind a black patch in *b*. When the white rectangle is behind, it 'captures' part of the black rectangle with it. This may be attributed to 'depth propagation'^{9,14}. *c*, Only the upper half of *a* is shown. It is very difficult to complete the monocular 'L'-shaped features and perceive them as black patches occluded by white patches. Yet vivid depth arises from this stereogram. *d*, If visual completion occurs to each monocular pattern, according to a recent 'reliability' theory, the completed part should be part of the ellipse. Therefore if the visually completed monocular patterns were the elements for stereoscopic depth under binocular fusion, then the curved boundaries from the completion would match with the straight real boundaries and produce an occluder that is cylindrical in shape. *e*, If we explicitly complete the ellipses, one can see the cylindrical occluder in front of a black ellipse. Comparison between the stereoscopic impressions of *d* and *e* leads us to reject the monocular completion explanation.

FIG. 3 A, Perspective view of limiting occlusion. The right edge of the white occluder in front and the right edge of the black patch are along the same visual direction in the left eye, therefore the right edge of the black patch is invisible to the left eye. Similarly, the left edge of the black patch is obscured from the right eye's view. B, Plan view of A shows only the section including the occluders. Black patch *pp* is pasted on a white background. Because of white occluder *aa*, only a strip of *pp* near its left edge is imaged to the left eye and only a strip near the right edge is imaged to the right eye. For a given bar width (represented by the darkly shaded areas), *aa* has the smallest distance from *pp*, because occluders closer to *pp* than *aa* cannot completely occlude *pp* on one side without changing the width of the vertical bars of the retinal images. But, *aa* is not the only occluder that produces these retinal images. Any white occluder that fits into the light grey area will produce vertical bars of the same width as *aa*. Occluder *bb* is larger than *aa* and is farther away from *pp*. Its right edge intrudes into the white background, but the resulting left retinal image is not different from that produced by



aa because the occluders are all assumed to have the same colour as the background. Therefore the situation is ambiguous. Despite the ambiguity inherent in this situation, observers usually perceive the occluder to be at position *aa*, as close as possible to the background (Fig. 2b).

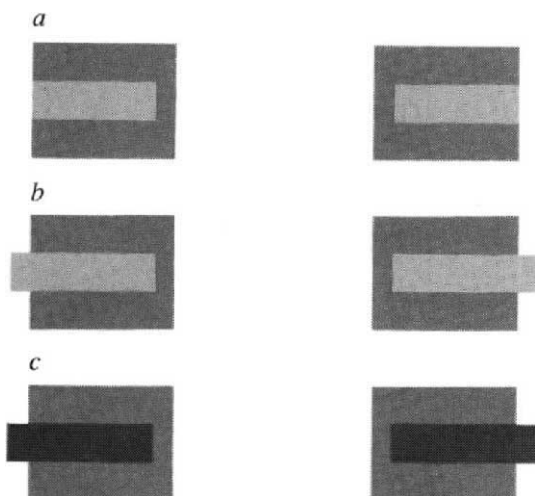


FIG. 4 Retinal images of occluders defined by contrast edges. *a*, The shade of occluder *aa* in Fig. 3B is light grey. *b*, The shade of occluder *bb* of Fig. 3B is light grey. *c*, The shade of occluder *bb* of Fig. 3B is black. Although *b* and *c* have the same geometric configuration and portray the same occluder, the stereoscopic impressions are very different. The occluder in the middle has less depth than the one below and a rivalrous region is attached to either end of the occluder. The perceptual difference can be attributed to the contrast polarities of matching edges. Consider the contrast polarities of the left edges of the occluders in the left half images. In *c*, it is light on the left and dark on the right. This edge will readily match with the left edge of the occluders in the right half image because the polarity there is also light on the left and dark on the right. Therefore the corresponding edges support occluder *bb* in Fig. 3B as a solution. In the left-hand image of *b*, however, it is also light on the left and dark on the right but the polarity of its corresponding edge in the right half image is dark on the left and light on the right. Because extended patches of opposite contrasts usually do not produce stereopsis¹¹, the corresponding edges in *b* do not support occluder *bb* in Fig. 3B and the visual system has to interpret them differently. In *a*, the occluder is at position *aa* (Fig. 3B), which matches the default assumption for depth based on occlusion, and so the percept is coincidentally veridical. Thus we show that in special situations, such as that in *a*, stimuli with opposite contrast polarities can give rise to vivid depth.

direction of the right eye (occluder *aa* in Fig. 3B).

Stereograms in Fig. 4*a-c* represent three real occlusions. When cross-fused, the apparent depth of the middle occluder is much less than the one below. One can see a lustrous fin attached to either side of the middle occluder, a sign of binocular rivalry. In Fig. 4*c*, the corresponding edges of the occluder have the same contrast polarity, therefore they support an occluder, which intrudes into the white background (*bb* in Fig. 3B). In Fig. 4*b*, the corresponding edges of the occluder have opposite contrast polarities, therefore they cannot support occluder *bb* because edges of opposite contrast polarities usually do not produce depth perception¹¹. In this case, even with the presence of monocular edges of occluder *bb*, our perception more readily accepts occluder *aa* as a solution, at the cost of throwing parts of occluder *bb* into rivalry. This is another situation where the same limiting occlusion is accepted as a default solution when there is no strong binocular support for a particular occluder. It appears that the binocularly fused vertical edges above and below the openings force this solution in both cases.

Current computational models for human stereopsis are based on the matching of corresponding luminance edges in the two eyes' images^{1,2}. Physiological studies have also found neurons in the visual cortex which respond to disparity of luminance edges^{12,13}. We have shown in this communication that under certain conditions, quantitative stereopsis arises even though the luminance edge exists only in one eye. We also demonstrated that under physically realizable conditions (Fig. 4*a*), stimuli with luminance edges of opposite contrast polarities may give rise to vivid depth. Our findings suggest that models for human stereopsis should treat features lacking binocular correspondence as indicators of occluding edges and decide whether the limiting occlusion should be applied to them. □

Received 19 April; accepted 4 October 1993.

1. Julesz, B. *Foundations of Cyclopean Perception* (University of Chicago Press, Chicago, London, 1971).
2. Marr, D. & Poggio, T. A. *Proc. R. Soc. B* **204**, 301-328 (1979).
3. Shimojo, S. & Nakayama, K. *Vision Res.* **30**, 69-80 (1990).
4. Nakayama, K. & Shimojo, S. *Vision Res.* **30**, 1811-1825 (1990).
5. Nakayama, K. & Shimojo, S. *Cold Spring Harb Symp. quant. Biol.* **55**, 911-924 (1990).
6. Harris, C. S. & Gregory, R. L. *Perception* **2**, 235-247 (1973).
7. Ramachandran, V. S. *Percept. Psychophys.* **39**, 361-373 (1986).
8. Ramachandran, V. S. & Cavanagh, P. *Nature* **317**, 527-530 (1985).
9. Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. *Nature* **242**, 412-414 (1973).

10. Kellman, P. J. & Shipley, T. F. *Cognitive Psychology* **23**, 141–221 (1991).
11. Treisman, A. Q. J. *exp. Psychol.* **14**, 23–37 (1962).
12. Barlow, H. B., Blakemore, C. & Pettigrew, J. D. *J. Physiol., Lond.* **193**, 327–342 (1967).
13. Poggio, G. F. & Fischer, B. *J. Neurophysiol.* **40**, 1392–1407 (1977).
14. Takeichi, H., Watanabe, T. & Shimizu, S. *Perception* **21**, 177–184 (1992).

ACKNOWLEDGEMENTS. We thank L. Cormack, L. Kontsevich and C. W. Tyler for helpful discussions. This study was supported by an NIH grant to C.M.S.

Active propagation of somatic action potentials into neocortical pyramidal cell dendrites

Greg J. Stuart & Bert Sakmann

Max-Planck-Institut für medizinische Forschung,
Abteilung Zellphysiologie, Postfach 103820,
69028 Heidelberg, Germany

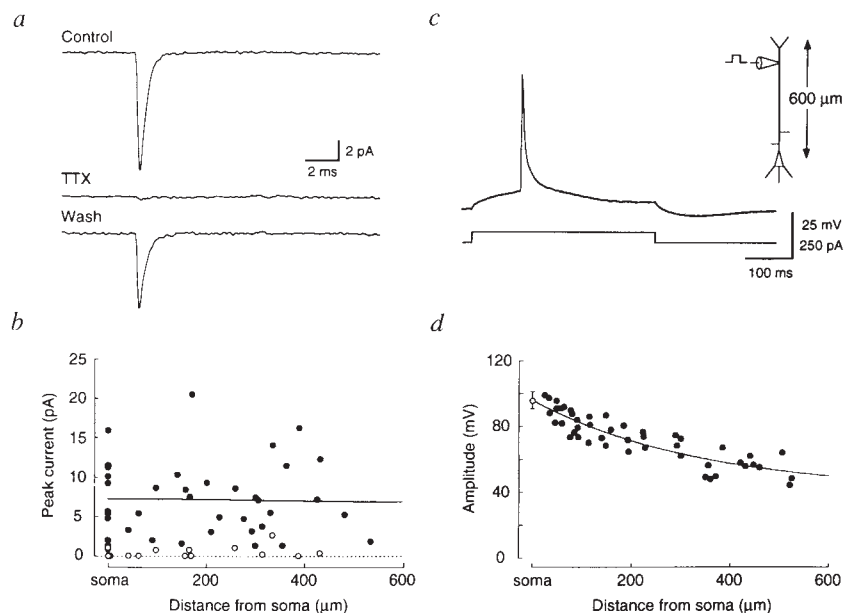
THE dendrites of neurons in the mammalian central nervous system have been considered as electrically passive structures which funnel synaptic potentials to the soma and axon initial segment, the site of action potential initiation^{1,2}. More recent studies, however, have shown that the dendrites of many neurons are not passive, but contain active conductances^{3,4}. The role of these dendritic voltage-activated channels in the initiation of action potentials in neurons is largely unknown. To assess this directly, patch-clamp recordings

were made from the dendrites of neocortical pyramidal cells in brain slices. Voltage-activated sodium currents were observed in dendritic outside-out patches, while action potentials could be evoked by depolarizing current pulses or by synaptic stimulation during dendritic whole-cell recordings. To determine the site of initiation of these action potentials, simultaneous whole-cell recordings were made from the soma and the apical dendrite or axon of the same cell. These experiments showed that action potentials are initiated first in the axon and then actively propagate back into the dendritic tree.

Patch-clamp recordings from the apical dendrite of layer V pyramidal neurons revealed that the dendritic membrane of these cells contains voltage-activated channels. In response to step depolarizations from a negative holding potential, rapidly inactivating inward currents were observed in outside-out patches excised up to 530 μm from the soma. As these inward currents were blocked by tetrodotoxin (TTX), they presumably represent the activation of dendritic voltage-activated Na^+ channels (Fig. 1a). Despite considerable variability, the peak Na^+ current in dendritic patches was not significantly different from that at the soma (Fig. 1b). On average, the peak Na^+ current was 7.7 ± 1.5 pA ($\pm$ s.e.m., $n=10$) in somatic patches and 7.0 ± 0.9 pA ($n=28$) in dendritic patches (excised 63–530 μm from the soma). This gives a somatic and dendritic Na^+ channel density of ~ 40 pS μm^{-2} (assuming a patch membrane area of ~ 1.5 μm^2 ; ref. 5). The finding that the apical dendrite of pyramidal neurons contains voltage-activated Na^+ channels agrees with previous work in both neocortical^{6–8} and hippocampal pyramidal neurons^{9–13}. In contrast to hippocampal pyramidal neurons,

FIG. 1 Voltage-activated Na^+ currents and action potentials recorded from the apical dendrite of neocortical layer V pyramidal neurons. **a**, Rapidly inactivating inward current evoked in an outside-out dendritic patch excised from the apical dendrite of a layer V pyramidal neuron (439 μm from the soma). This current was blocked in the presence of 500 nM TTX. **b**, Plot of the peak inward current evoked in somatic ($n=10$) and dendritic ($n=28$) outside-out patches by depolarizations to the same test potential (filled symbols). The data have been fitted by a linear regression (solid line; regression coefficient is 0.028). When tested, this inward current was substantially reduced by 100 or 500 nM TTX (open symbols; $n=14$). **c**, Dendritic action potential evoked at the resting membrane potential (-64 mV) by a 60 pA depolarizing current pulse during whole-cell recording from the apical dendrite of a layer V pyramidal neuron (440 μm from the soma). The inset in this and the following figures shows a schematic representation of the experimental recording arrangement. **d**, Plot of dendritic action potential amplitude (filled symbols) at different distances from the soma ($n=39$; some cells were patched twice at different dendritic locations). The average amplitude ($\pm$ s.d.) of somatic action potentials recorded in 16 of these cells is indicated (open symbol). The data have been arbitrarily fitted by a single exponential.

METHODS. Experiments were performed on 300- μm -thick parasagittal neocortical brain slices from 2-week-old Wistar rats. Slices were perfused with an oxygenated Ringer solution containing (in mM): 125 NaCl, 25 NaHCO_3 , 25 glucose, 2.5 KCl, 1.25 NaH_2PO_4 , 2 CaCl_2 , 1 MgCl_2 (pH 7.4) and, except when otherwise stated, experiments were performed at room temperature. Patch-clamp recordings were made from visually identified apical dendrites and somata of layer V pyramidal neurons using previously described techniques^{29,30}. All outside-out patch-clamp recordings were made with a patch-clamp amplifier (EPC-7, List, Darmstadt, Germany) using 10-M Ω patch pipettes coated with Sylgard (184, Dow Corning) and filled with the following solution (in mM): 140 CsCl, 10 HEPES, 10 ethylene glycol tetraacetate (EGTA), 2 MgCl_2 , 2 $\text{Na}_2\text{-ATP}$ (pH 7.3). Rapidly inactivating inward currents were evoked



by depolarizing steps to -10 mV from a holding potential of -90 mV and leak and capacitive currents were subtracted on-line (each trace is the average of 10 sweeps). Whole-cell recordings were made with a microelectrode amplifier (Axoclamp 2A, Axon Instruments, Foster City, US) using 10–15-M Ω patch pipettes filled with the following solution (in mM): 120 K-gluconate, 20 KCl, 10 HEPES, 10 EGTA, 2 MgCl_2 , 2 $\text{Na}_2\text{-ATP}$ (pH 7.3). Action potentials were evoked by depolarizing current pulses from the resting membrane potential and their amplitude was measured from a baseline set at threshold. The location of dendritic recordings was measured directly from the video monitor using an image processor (Argus 10, Hamamatsu, Hamamatsu City, Japan) and gives the distance from the centre of the soma to the dendritic recording site. Recordings were confirmed to be dendritic (outside-out patch recordings, $n=4$; whole-cell recordings, $n=6$) by identifying the recording site under fluorescence optics following filling of the neuron with the fluorescent dye Lucifer yellow (Sigma)³⁰.

however, where Na^+ channels are thought to be restricted to within 200 μm of the soma^{12,13}, neocortical pyramidal neurons express Na^+ channels along the entire length of the apical dendrite (Fig. 1b).

To investigate the possibility that dendritic Na^+ channels can initiate action potentials in the dendrites, as has been suggested in both neocortical^{8,14–16} and hippocampal pyramidal cells^{9,17–20}, whole-cell voltage recordings were made from the apical dendrite of layer V pyramidal neurons. Action potentials could be evoked by depolarizing current pulses during recordings up to 525 μm from the soma (Fig. 1c). These action potentials, which were blocked by TTX (1 μM), were smaller and broader than somatic action potentials and their amplitude decreased with increasing distance from the soma (Fig. 1d). A similar decrease in dendritic action potential amplitude with increasing distance from the soma has also been observed in hippocampal pyramidal neurons²¹.

One explanation for these findings is that dendritic action

potentials represent an attenuated form of an action potential initiated close to the soma (where action potential amplitude is largest). To test this possibility, simultaneous whole-cell recordings were made from the soma and apical dendrite of the same cell. Action potentials evoked by dendritic current pulses (60–800 pA; 40–500 ms) were observed first at the soma (Fig. 2a; $n = 32$ at room temperature, dendritic recordings 60–525 μm from the soma; $n = 8$ at 35 °C, dendritic recordings 140–400 μm from the soma). That somatic and dendritic recordings were in fact made from the same cell, rather than electrically coupled cells²², was demonstrated by simultaneously filling the same cell from the soma and the dendrite with different coloured fluorescent dyes (Fig. 2b; $n = 20$). Large-amplitude (up to 15 nA) dendritic current pulses of short duration (0.5 ms) were used in an attempt to initiate action potentials in the dendrites ($n = 5$ at 35 °C, dendritic recordings 140–350 μm from the soma). At threshold, action potentials evoked by short dendritic current pulses occurred first at the soma. As the amplitude of these

FIG. 2 Site of dendritic action potential initiation in neocortical layer V pyramidal neurons. **a**, Simultaneous whole-cell recordings were made from the soma and apical dendrite of the same layer V pyramidal neuron (dendritic recording 525 μm from the soma). Action potentials were evoked by depolarizing current pulse to either the soma (top; 150 pA) or dendrite (bottom; 300 pA). **b**, Simultaneous filling of the same layer V pyramidal neuron from the dendrite and the soma with different coloured fluorescent dyes: Cascade blue at the soma and Lucifer yellow in the dendrite (dendritic recording 190 μm from the soma). **c**, Action potentials initiated by distal synaptic stimulation in layer I (arrow) during simultaneous somatic and dendritic recording (dendritic recording 525 μm from the soma; same cell as in **a**). **d**, Plot of the difference in latency of the peak of dendritic and somatic action potentials during simultaneous somatic and dendritic recordings at different distances from the soma. Action potentials were initiated by either somatic current pulses (filled symbols; $n = 32$) or distal synaptic stimulation (open symbols; $n = 5$). The data have been fitted by a linear regression (regression coefficient is 0.88).

METHODS. Simultaneous whole-cell recordings were made from the soma and dendrite of the same visually identified layer V pyramidal cell using two identical microelectrode amplifiers (Axoclamp 2A). Action potentials were evoked by depolarizing current pulses to either the somatic or dendritic patch pipette. Recordings were only made from cells where it was possible to follow the apical dendrite from the cell soma to the site of dendritic recording. Cascade blue (Molecular Probes, Eugene, US) and Lucifer yellow (Sigma) were added directly to the intracellular solution at a concentration of 1 and 4 mg ml^{-1} , respectively. Both dyes can be excited by light with a wavelength of less than 410 nm and their fluorescence viewed simultaneously using an appropriate filter set (excitation BP 365/12, dichroic FT 395, emission LP 397; Zeiss). For synaptic stimulation, a third patch pipette (outer diameter 10–20 μm , filled with Ringer solution) was positioned under visual control in layer I and e.p.s.ps were evoked by brief voltage pulses (200 μs , up to 50 V).

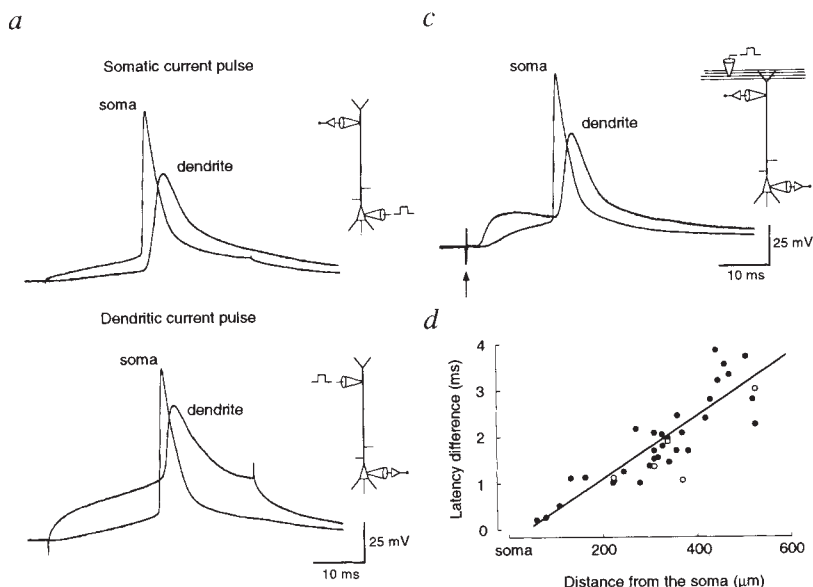
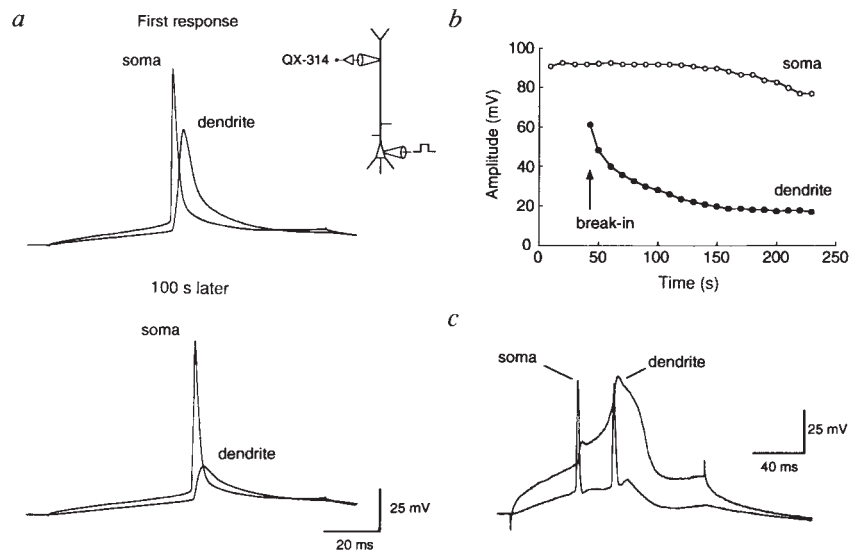


FIG. 3 Effect of the intracellular Na^+ channel blocker QX-314 on dendritic action potentials. **a**, Top, Somatic and dendritic action potentials, evoked by a depolarizing somatic current pulse (100 pA), during simultaneous somatic and dendritic recording immediately after obtaining a dendritic whole-cell recording (443 μm from the soma) with 10 mM QX-314 (Astra) in the dendritic patch pipette. Bottom, The same recording 100 s later shows that diffusion of QX-314 into the dendrite from the dendritic patch pipette substantially reduced the amplitude of the dendritic action potential with little effect on the action potential at the soma. **b**, The amplitude of the somatic (open symbols) and dendritic (filled symbols) action potentials for the experiment shown in **a** have been plotted against time. Break-in to obtain a dendritic whole-cell recording is indicated by the arrow. **c**, Dendritic Ca^{2+} spike evoked in isolation from the soma by a dendritic depolarizing current pulse (200 pA) during simultaneous somatic and dendritic recordings (dendritic recording 385 μm from the soma) with 1 mM QX-314 in the dendritic patch pipette.



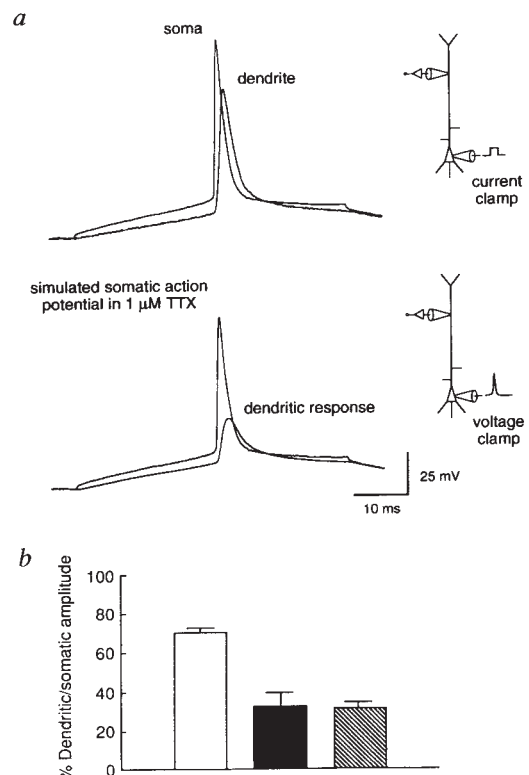
dendritic current pulses was increased, however, the time between the peaks of the dendritic and somatic action potentials decreased, and on one occasion (260 μm from the soma, 5 nA current pulse) the dendritic action potential was observed to come before the somatic action potential. These results show that although action potential initiation usually occurs at or near the soma, it can shift into the dendrites during very strong dendritic depolarizations (to ~ 0 mV). To determine whether this can occur during synaptic stimulation, as has recently been suggested⁸, action potentials were initiated by excitatory synaptic potentials (e.p.s.ps). Stimulation of excitatory synapses in layer I of the neocortex was used to depolarize the very distal regions of the dendritic tree. Action potentials initiated by these distal e.p.s.ps, which were larger and faster in the dendrite, were always observed to occur first at the soma (Fig. 2c; $n=5$ at room temperature; $n=8$ at 35 $^{\circ}\text{C}$). Similar results were also obtained when

e.p.s.ps were evoked by extracellular stimulation in other layers of the neocortex ($n=11$) or in the presence of bicuculline to block inhibition (5 μM). We conclude that, during synaptic stimulation of layer V pyramidal neurons, action potential initiation is at or near the soma.

To establish whether action potential initiation was somatic or axonal in origin, simultaneous recordings were obtained from the soma and axon of the same cell. Action potentials initiated by either somatic or axonal current pulses were always observed first in the axon ($n=4$; axonal recordings 10–35 μm from the soma). This provides direct evidence for the hypothesis that action potential initiation in neurons occurs in the axon, possibly at the axon initial segment, as suggested in the original studies on action potential initiation in motor neurons¹. The finding that action potential initiation occurs first in the axon, rather than in the soma or dendrites, is presumably due to differences

FIG. 4 Comparison of active and passive propagation of evoked action potentials and simulated action potential waveforms. **a**, Top, Somatic and dendritic action potentials evoked by a depolarizing somatic current pulse (200 pA) during simultaneous somatic and dendritic recording (dendritic recording 310 μm from the soma). Bottom, The dendritic response (at the same dendritic location) to a simulated somatic action potential waveform in the presence of TTX together with the actual voltage change at the soma during this simulated somatic action potential. **b**, Comparison of the average amplitude of dendritic action potentials (open column) and dendritic responses to inverted (hatched column) or upright (solid column) simulated somatic action potentials recorded at the same dendritic location (data expressed as a percentage of the response recorded at the soma $\pm$ s.e.m.; dendritic recordings 165–470 μm from the soma).

METHODS. Simultaneous whole-cell recordings were made using a patch-clamp amplifier (EPC-7) at the soma and a microelectrode amplifier (Axoclamp 2A) at the dendrite. Somatic and dendritic action potentials were evoked by a depolarizing somatic current pulse. The slice was then perfused with TTX (1 μM) and the previously recorded somatic action potential was used as a voltage-clamp command at the soma during somatic whole-cell voltage clamp. The dendritic response to this simulated action potential was then recorded at the same dendritic location. The actual voltage change at the soma during the simulated action potential was later recorded at the soma with a second somatic pipette. Somatic voltage-clamp patch pipettes (2–3 M Ω) were coated with Sylgard and series resistance (7–10 M Ω) was compensated by 70–90%. The action potential voltage command during somatic voltage clamp was scaled to one-tenth of its original size and inverted ($n=7$) or was the original size and upright in which case the experiments were performed in the presence of TTX ($n=5$). On three occasions both inverted and upright simulated action potentials were used in the same cells.



in the geometry, as well as to possible differences in the density, distribution and properties of voltage-activated channels in these structures²³.

The difference in latency of the peak of dendritic and somatic action potentials, evoked by somatic depolarization or synaptic stimulation, indicates a soma-to-dendrite action potential conduction velocity of $\sim 0.15 \text{ m s}^{-1}$ at room temperature (Fig. 2d), that is, the peak of an action potential takes 3–4 ms to propagate from the soma to the most distal apical dendrites. At 35 °C this conduction velocity was two to three times faster.

To investigate whether dendritic voltage-activated Na^+ channels aid the back-propagation of somatic action potentials into the dendrites, the internal Na^+ channel blocker QX-314 (ref. 24) was included in the dendritic patch pipette during simultaneous somatic and dendritic recording. Following establishment of dendritic recording, dendritic action potentials were observed to decrease progressively in amplitude before any change was observed in the amplitude or time course of the somatic action potential (Fig. 3a, b; $n=8$). This suggests that dendritic Na^+ channels are not involved in the initiation of somatic action potentials, but boost the amplitude of dendritic action potentials as they propagate back into the dendritic tree. In the presence of QX-314, dendritic Ca^{2+} spikes (blocked by 200 μM cadmium chloride) could be evoked in isolation from the soma (Fig. 3c), showing that, in addition to Na^+ channels, the dendrites of layer V pyramidal neurons also contain voltage-activated Ca^{2+} channels (see also ref. 7).

To compare the expected attenuation of dendritic action potentials in the presence and absence of dendritic Na^+ channels, voltage commands were applied at the soma to simulate a somatic action potential (in the presence of TTX or inverted) and the amount of attenuation of this simulated action potential waveform was compared with that of evoked action potentials. Somatic and dendritic action potentials were first recorded in response to a somatic current pulse (Fig. 4a; top). Following addition of 1 μM TTX, the same action potential as previously recorded at the soma was used as the voltage command during somatic voltage clamp and the response to this simulated action potential recorded at the same dendritic location (Fig. 4a; bottom). On average, for these experiments the amplitude of evoked dendritic action potentials was $70 \pm 2.6\%$ ($\pm \text{s.e.m.}$, $n=9$) of somatic action potentials (during dendritic recordings 165–470 μm from the soma), whereas the dendritic response recorded at the same location to either a simulated upright action potential in TTX or an inverted action potential waveform (scaled to one-tenth the size of the somatic action potential) was $33 \pm 6.7\%$ ($n=5$) and $32 \pm 3.3\%$ ($n=7$) of that recorded at the soma (Fig. 4b). These results show unequivocally that somatic action potentials actively invade the dendrites of layer V pyramidal neurons via the activation of dendritic TTX-sensitive Na^+ channels.

Back-propagation of somatic action potentials provides neurons with a rapid retrograde signal that could modulate the computational properties of neurons in both the short and long term. For example, in the short term, dendritic action potentials may activate dendritic voltage-activated Ca^{2+} channels (Fig. 3c), leading to a rise in dendritic intracellular Ca^{2+} that could transiently shunt out parts of the dendritic tree by opening Ca^{2+} -dependent conductances. Sodium entry itself may have the same effect by opening dendritic Na^+ -activated K^+ conductances. Active propagation of somatic action potentials into the dendritic tree may also be important for the induction of long-term changes in synaptic strength, either by direct activation of dendritic voltage-activated Ca^{2+} channels^{13,25,28} or by boosting synaptic Ca^{2+} influx by briefly relieving the voltage-dependent Mg^{2+} block of dendritic *N*-methyl-D-aspartic acid receptor channels. □

- Rall, W. in *Handbook of Physiology. The Nervous System* (ed. Kandel, E. R.) 39–97 (Am. Physiol. Soc., Bethesda, Maryland, 1977).
- Llinas, R. R. *Science* **242**, 1654–1664 (1988).
- Adams, P. R. *Curr. Biol.* **2**, 625–627 (1992).
- Sakmann, B. & Neher, E. in *Single-Channel Recording* (eds Sakmann, B. & Neher, E.) 37–51 (Plenum, New York, 1983).
- Huguenard, J. R., Hamill, O. P. & Prince, D. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2473–2477 (1989).
- Amitai, Y., Friedman, A., Connors, B. W. & Gutnick, M. J. *Cerebral Cortex* **3**, 26–38 (1993).
- Regehr, W., Kehoe, J., Ascher, P. & Armstrong, C. *Neuron* **11**, 145–151 (1993).
- Wong, R. K. S., Prince, D. A. & Basbaum, A. I. *Proc. natn. Acad. Sci. U.S.A.* **76**, 986–990 (1979).
- Benardo, L. S., Masukawa, L. M. & Prince, D. A. *J. Neurosci.* **2**, 1614–1622 (1982).
- Miyakawa, H. & Kato, H. *Brain Res.* **399**, 303–309 (1986).
- Turner, R. W., Meyers, D. E. R. & Barker, J. L. J. *Neurophysiol.* **62**, 1375–1387 (1989).
- Jaffe, D. B. et al. *Nature* **357**, 244–246 (1992).
- Purpura, D. P. in *The Neurosciences* (eds Quarten, G. C., Melnechuk, T. & Schmitt, F. O.) 372–393 (Rockefeller Univ. Press, New York, 1967).
- Deschenes, M. *Expl Brain Res.* **43**, 304–308 (1981).
- Pockberger, H. *Brain Res.* **539**, 181–190 (1991).
- Spencer, W. A. & Kandel, E. R. *J. Neurophysiol.* **24**, 272–285 (1961).
- Anderson, P. & Lomo, T. *Expl Brain Res.* **2**, 247–260 (1966).
- Anderson, P., Storm, J. & Wheal, H. V. *J. Physiol., Lond.* **383**, 509–526 (1987).
- Herreras, O. *J. Neurophysiol.* **64**, 1429–1441 (1990).
- Turner, R. W., Meyers, D. E. R., Richardson, T. L. & Barker, J. L. J. *J. Neurosci.* **11**, 2270–2280 (1991).
- Peinado, A., Yuste, R. & Katz, L. C. *Neuron* **10**, 103–114 (1993).
- Moore, J. W., Stockbridge, N. & Westerfield, M. J. *Physiol., Lond.* **336**, 301–311 (1983).
- Strichartz, G. R. *J. gen. Physiol.* **62**, 37–57 (1973).
- Artola, A., Bröcher, S. & Singer, W. *Nature* **347**, 69–72 (1990).
- Grover, L. M. & Teyler, T. J. *Nature* **347**, 477–479 (1990).
- Jaffe, D. & Johnston, D. *J. Neurophysiol.* **64**, 948–960 (1990).
- Komatsu, Y. & Iwakiri, M. *J. Neurophysiol.* **67**, 401–410 (1992).
- Dodt, H.-U. & Ziegglansberger, W. *Brain Res.* **537**, 333–336 (1990).
- Stuart, G. J., Dodt, H.-U. & Sakmann, B. *Pflügers Arch. ges. Physiol.* **423**, 511–518 (1993).

ACKNOWLEDGEMENTS. We thank M. Häusser, N. Spruston and D. Johnston for discussions, F. Helmchen for computer programming and M. Häusser and N. Spruston for comments on the manuscript. G.J.S. acknowledges funding by a C. J. Martin scholarship from the National Health and Medical Research Council of Australia.

Three types of Ca^{2+} channel trigger secretion with different efficacies in chromaffin cells

Cristina R. Artalejo^{*†}, Michael E. Adams[‡] & Aaron P. Fox[§]

^{*} Northwestern University, Department of Neurobiology and Physiology, 2153 North Campus Drive, Evanston, Illinois 60208, USA
[†] Departamento de Farmacología, Facultad de Medicina, Universidad Autónoma de Madrid, Madrid 28029, Spain
[‡] University of California at Riverside, Department of Entomology and Neuroscience, Riverside, California 92521, USA
[§] Department of Pharmacological and Physiological Sciences, University of Chicago, 947 East 58th Street, Chicago, Illinois 60637, USA

To determine whether the different types of Ca^{2+} channels present in the same secretory cell contribute equally to secretion, we used chromaffin cells to analyse the coupling between three distinct types of Ca^{2+} channel¹ and exocytosis^{2,3}. These are ω -conotoxin-GVIA-sensitive^{4,5} N-type channels, ω -agatoxin-IVA-sensitive P-type Ca^{2+} channels⁶ and dihydropyridine-sensitive facilitation Ca^{2+} channels, which are normally quiescent but are activated by depolarizing pre-pulses, repetitive depolarizations to physiological potentials^{7,8}, or agents that raise cyclic AMP⁹. We have simultaneously monitored changes in capacitance as an assay of catecholamine secretion¹⁰, and Ca^{2+} currents. Although all three types of Ca^{2+} channel trigger secretion individually, facilitation channels produce much greater secretion for a given size of Ca^{2+} current, indicating that they are coupled more efficiently to exocytosis. These results indicate that facilitation Ca^{2+} channels may be physically nearer vesicle release sites. They also show that low efficiency P- and N-type channels could trigger mild release and that high-efficiency facilitation channels may underlie the massive catecholamine release that occurs during the 'fight or flight' response.

TABLE 1 Ca^{2+} channel-secretion statistics

Ca^{2+} channel type	Total capacitance change (fF)	Total number of vesicles*	Peak current (pA)	Current integral (Ca^{2+} ions)	Vesicles per Ca^{2+} ion $\times 10^{-6}$	Rate (fF ms^{-1})
Facilitation ($n=16$)	579.09 $\pm$ 44.27	289.55 $\pm$ 22.135	188.318 $\pm$ 6.769	23.019 $\pm$ 1.485	14.578 $\pm$ 1.5101	0.27839 $\pm$ 0.0082
P-type ($n=21$)	165.166 $\pm$ 7.486	82.58 $\pm$ 3.74	237.25 $\pm$ 15.6	32.48 $\pm$ 2.028	2.9258 $\pm$ 0.2318	0.09203 $\pm$ 0.0052
N-type ($n=17$)	161.75 $\pm$ 9.022	80.88 $\pm$ 4.511	253.25 $\pm$ 16.63	33.03 $\pm$ 2.28	2.8542 $\pm$ 0.2749	0.0933 $\pm$ 0.005

* One vesicle corresponds to 2fF.

The combination of nisoldipine, ω -conotoxin-GVIA (ω -CgTx) and ω -agatoxin-IVA (ω -Aga-IVA) completely suppressed Ca^{2+} currents in chromaffin cells (Fig. 1). The current recorded in the presence of 5 μM ω -CgTx was $\sim 60\%$ of that recorded under control conditions (Fig. 1A, a, traces 1 and 2). Nisoldipine (1 μM) did not affect the residual current, indicating

that facilitation Ca^{2+} current was not part of this component (trace 3). ω -Aga-IVA (100 nM) completely blocked the residual current (trace 4). Trains of depolarizations to +140 mV reversed the ω -Aga-IVA¹¹ block (trace 5). Figure 1A, b shows the same data plotted as peak current versus time. Together with previous data^{1,7,8} these results indicate that chromaffin cells have three

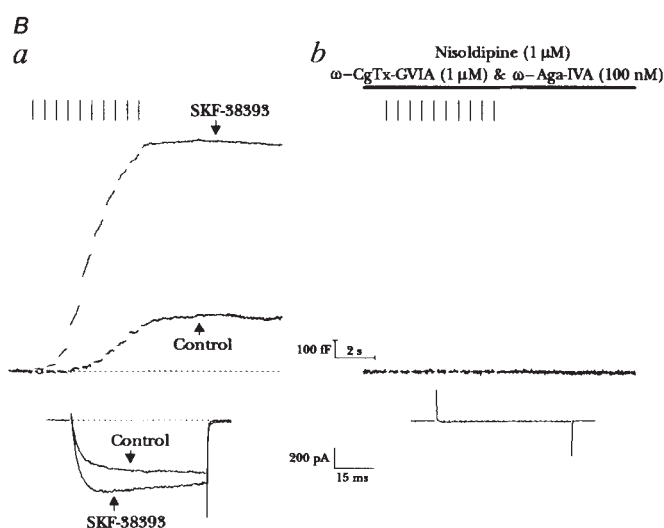
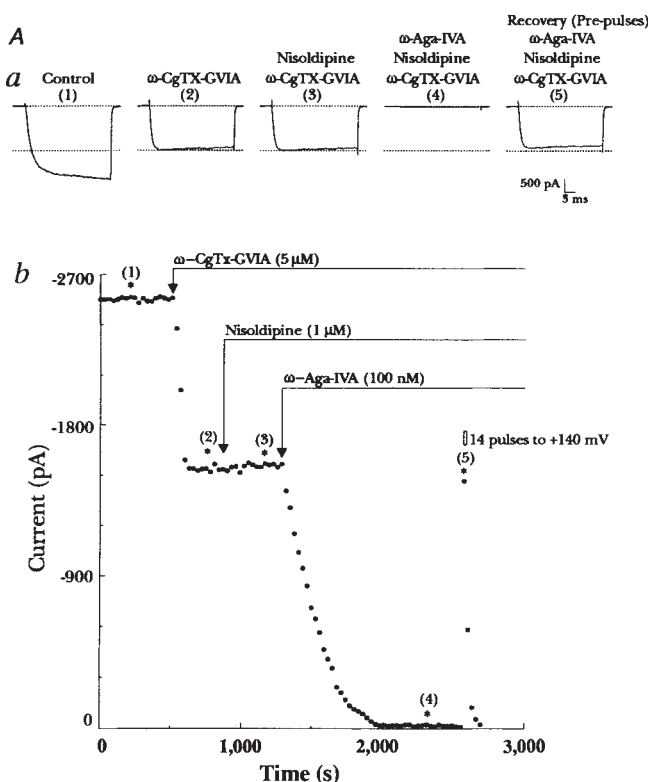


FIG. 1 A, Chromaffin cell Ca^{2+} currents are completely suppressed by a combination of different antagonists. a, Leak-substituted Ca^{2+} current traces obtained by depolarizing a chromaffin cell to +10 mV from a holding potential of -80 mV. Labels on top of current records indicate the antagonists present during the recordings. Panel 1 shows control Ca^{2+} current (facilitation Ca^{2+} current was not activated). Application of ω -CgTx (5 μM) decreased the current by 40% (panel 2). Nisoldipine had no effect as facilitation Ca^{2+} currents were not activated (panel 3). ω -Aga-IVA (100 nM) blocked all the remaining current (panel 4). A train of 14 depolarizations lasting 50 ms from a holding potential of -80 mV to +140 mV (repeated at 1-s intervals) reversed the blockade by ω -Aga-IVA (panel 5). On average, ω -Aga-IVA reduced the initial control current by $53 \pm 4.1\%$ ($n=13$). b, Plot of peak current versus time in absence and presence of Ca^{2+} channel antagonists, for data shown in a. The cell was depolarized to +10 mV from a holding potential of -80 mV once every 30 s. Compounds were added as indicated. Currents shown in A were obtained at the points labelled with asterisks in the graph. Similar results were obtained in a total of 9 experiments. B, recruiting facilitation Ca^{2+} channels dramatically augments secretion from chromaffin cells. a, Changes in membrane capacitance elicited by a train of 10 depolarizations to +10 mV, each lasting 50 ms (holding potential -90 mV); 500 ms separated each depolarization. Breaks in the capacitance records correspond to application of test depolariza-

tions. Identical trains were delivered to the same cell in drug-free conditions (control) or in the presence of the D_1 dopamine receptor agonist SKF-38393 (10 μM) to recruit facilitation Ca^{2+} current; 10 min separated each train. The corresponding Ca^{2+} currents elicited by the first depolarization of each train in the absence (control) or in the presence of SKF-38393 are plotted underneath. The 500 pA (64.68×10^6 Ca^{2+} ions) of P- plus N-type Ca^{2+} current (control) produced a 280 fF increase in the capacitance; in contrast, the 190 pA (24.50×10^6 Ca^{2+} ions) of facilitation Ca^{2+} current recruited by the D_1 agonist elicited an 800 fF increase in capacitance. In other experiments where facilitation currents were not activated, delivery of 4 identical trains produced 4 similar changes in capacitance. Thus the potentiation of secretion during the second (SKF-38393) train seen here is due exclusively to the recruitment of facilitation Ca^{2+} current and not to an effect of Ca^{2+} influx from the control train. b, Addition of nisoldipine (1 μM), ω -CgTx (1 μM) and ω -Aga-IVA (100 nM) completely suppressed Ca^{2+} currents as well as the changes in capacitance. Similar results were obtained in a total of 18 experiments.

METHODS. A, The pipette solution consisted of (mM): CsCl (100), Cs-EGTA (10), HEPES (40), MgCl_2 (5), ATP (2), GTP (0.3), adjusted to pH 7.2 with CsOH. The external solution consisted of (mM): BaCl_2 (10), tetraethylammonium (TEA)-Cl (150), HEPES (10) and 1 μM tetrodotoxin (TTX), adjusted to pH 7.2 with TEA-OH. Blockers were applied through a fast-perfusion system. Patch-clamp recording conditions were similar to those described⁵. Similar results were obtained using CaCl_2 as charge carrier. ω -Aga-IVA was isolated from the venom of the funnel web spider *Agelenopsis aperta* as described⁶. B, The pipette solution consisted of (mM): caesium glutamate (110), Cs-EGTA (0.1), HEPES (40), MgCl_2 (5), ATP (2), GTP (0.3), adjusted to pH 7.2 with CsOH. The external solution consisted of (in mM): CaCl_2 (2), TEA-Cl (150), HEPES (10) and 1 μM TTX, adjusted to pH 7.2 with TEA-OH. Capacitance was measured by a computer program using a phase-tracking technique^{20,21}.

pharmacologically distinct types of Ca^{2+} channel and that individual components can be selectively blocked using the appropriate antagonist. With this information it became possible to test the role of each Ca^{2+} channel in catecholamine secretion.

Recruitment of facilitation Ca^{2+} current greatly potentiated secretion. Adding a D_1 dopamine receptor agonist (SKF-38393) to the bath recruits facilitation current by a cAMP/protein kinase A cascade⁹. Figure 1*B, a* shows the capacitance change elicited by two identical trains of 10 depolarizations to +10 mV, before (smaller trace) or after (larger trace) recruiting facilitation current. The respective Ca^{2+} current traces are shown underneath. Under these conditions, no inactivation of the Ca^{2+} currents occurs along the 10 depolarizations of the train. The combination of nisoldipine (1 μM), ω -CgTx (1 μM) and ω -Aga-IVA (100 nM) inhibited all of the Ca^{2+} currents and prevented any change in capacitance (Fig. 1*B, b*). The results shown in Fig. 1*B* reveal that all the Ca^{2+} channels found in chromaffin cells may contribute to secretion, but that activating facilitation Ca^{2+} currents dramatically enhances catecholamine secretion.

Combinations of antagonists were used to measure the

secretion produced by individual Ca^{2+} channels. The combination of ω -CgTx and ω -Aga-IVA was used to suppress non-facilitation components (Fig. 2*A, a*). A train of depolarizations was applied after treatment with a D_1 agonist to recruit facilitation current. The Ca^{2+} currents obtained during the first depolarization of the train are shown below. Addition of nisoldipine (1 μM) completely inhibited facilitation current and the associated increase in capacitance (Fig. 2*A, b*). P-type Ca^{2+} currents also promoted catecholamine secretion, but did so less efficiently than facilitation Ca^{2+} currents (Fig. 2*B, a*). Nisoldipine and ω -CgTx were used to isolate P-type Ca^{2+} currents. ω -Aga-IVA blocked the P-type current and the associated secretion (Fig. 2*B, b*). N-type Ca^{2+} currents, isolated using nisoldipine and ω -Aga-IVA, were also found to be coupled to secretion with lower efficacy than facilitation Ca^{2+} currents (Fig. 2*C, a*). Adding ω -CgTx blocked the N-type current and the capacitance changes (Fig. 2*C, b*). A summary of 54 similar experiments is shown in Table 1. Figure 2 confirms that all three types of Ca^{2+} channel found in chromaffin cells promote secretion, but that facilitation Ca^{2+} channels do so more efficiently.

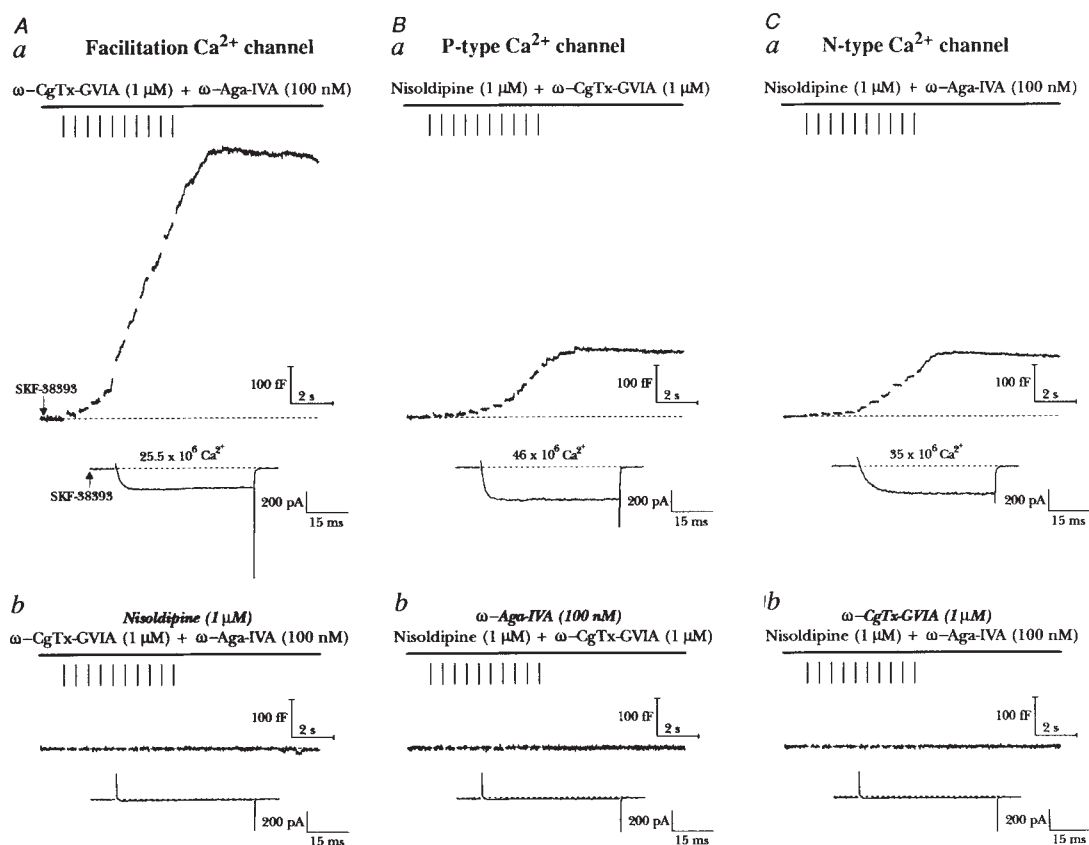


FIG. 2 Pharmacological dissection of the contribution of individual Ca^{2+} channel types to secretion: facilitation Ca^{2+} channels are the most efficient. *A, a*, Capacitance signal induced by activation of facilitation Ca^{2+} channels. A train of 10 depolarizations, similar to that described for Fig. 1*B*, was applied in the presence of a D_1 agonist to recruit facilitation current. Only facilitation Ca^{2+} channels were available; ω -CgTx and ω -Aga-IVA were present to suppress the P- and N-type Ca^{2+} channels. The Ca^{2+} current elicited by the first test depolarization is plotted underneath the capacitance trace. *b*, Addition of nisoldipine completely suppressed the Ca^{2+} currents and changes in the capacitance. *B, a*, Nisoldipine and ω -CgTx were used to suppress facilitation and N-type Ca^{2+} channels, leaving only the P-type channel unblocked. *b*, Addition of ω -Aga-IVA blocked the P-type Ca^{2+} current and associated capacitance change. *C, a*, Nisoldipine and ω -Aga-IVA suppressed the facilitation- and P-type Ca^{2+} channels, leaving only N-type channels

unblocked. *b*, Addition of ω -CgTx blocked the N-type current and changes in capacitance. The number at the top of the current trace corresponds to the number of Ca^{2+} ions entering the cell during the 50-ms depolarization. Although *A, B* and *C* are recordings from three different cells, they are representative of data from 54 chromaffin cells. For equivalent Ca^{2+} current amplitudes (facilitation: 190 pA (25.5 $\times 10^6 \text{ Ca}^{2+}$ ions); P-type: 320 pA (46 $\times 10^6 \text{ Ca}^{2+}$ ions); N-type: 260 pA (35 $\times 10^6 \text{ Ca}^{2+}$ ions)), the recruitment of facilitation Ca^{2+} channels resulted in the exocytosis of ~356 vesicles whereas activation of P- or N-type Ca^{2+} channels only resulted in the exocytosis of ~88 or ~83 vesicles respectively. Similar results were obtained when the individual Ca^{2+} current components were sequentially activated and later blocked in the same cell. In our hands, trains of longer depolarizations were less reproducible. All our capacitance experiments use relatively short depolarizations to trigger secretion.

To test whether increasing the internal concentration of cAMP ($[cAMP]_i$) could alter secretion by a mechanism not directly involving facilitation Ca^{2+} channels, P-type-channel-mediated secretion was measured in the presence and absence of the membrane-permanent cAMP analogue 8-chlorophenylthio-cAMP (8-cpt-cAMP) (Fig. 3A, *a, b*). Two identical trains of depolarizations produced virtually the same capacitance change, indicating that elevating $[cAMP]_i$ was without effect. Furthermore, recruiting facilitation with pre-pulses to +120 mV elicited capacitance changes identical to those produced when either D_1 agonists or 8-cpt-cAMP were used to recruit facilitation (Fig. 3B, *a, b*). When depolarizations were not applied, elevated $[cAMP]_i$ by

itself did not elicit capacitance changes (Fig. 3C, *a, b*). Thus our data strongly suggest that the result of activating D_1 receptors is due exclusively to recruitment of facilitation channels and not to a cAMP/protein kinase A effect on the secretory process downstream from Ca^{2+} channel activation¹². A recent study found that elevated $[cAMP]_i$ increased catecholamine secretion from chromaffin cells independently of Ca^{2+} channels¹³. The apparent discrepancy may be due to the fact that our experiments measure secretion on a millisecond timescale and probably correspond to docked or nearly releasable vesicles¹⁴, whereas the other study measures secretion after 15 min, when other physiological processes may contribute to secretion.

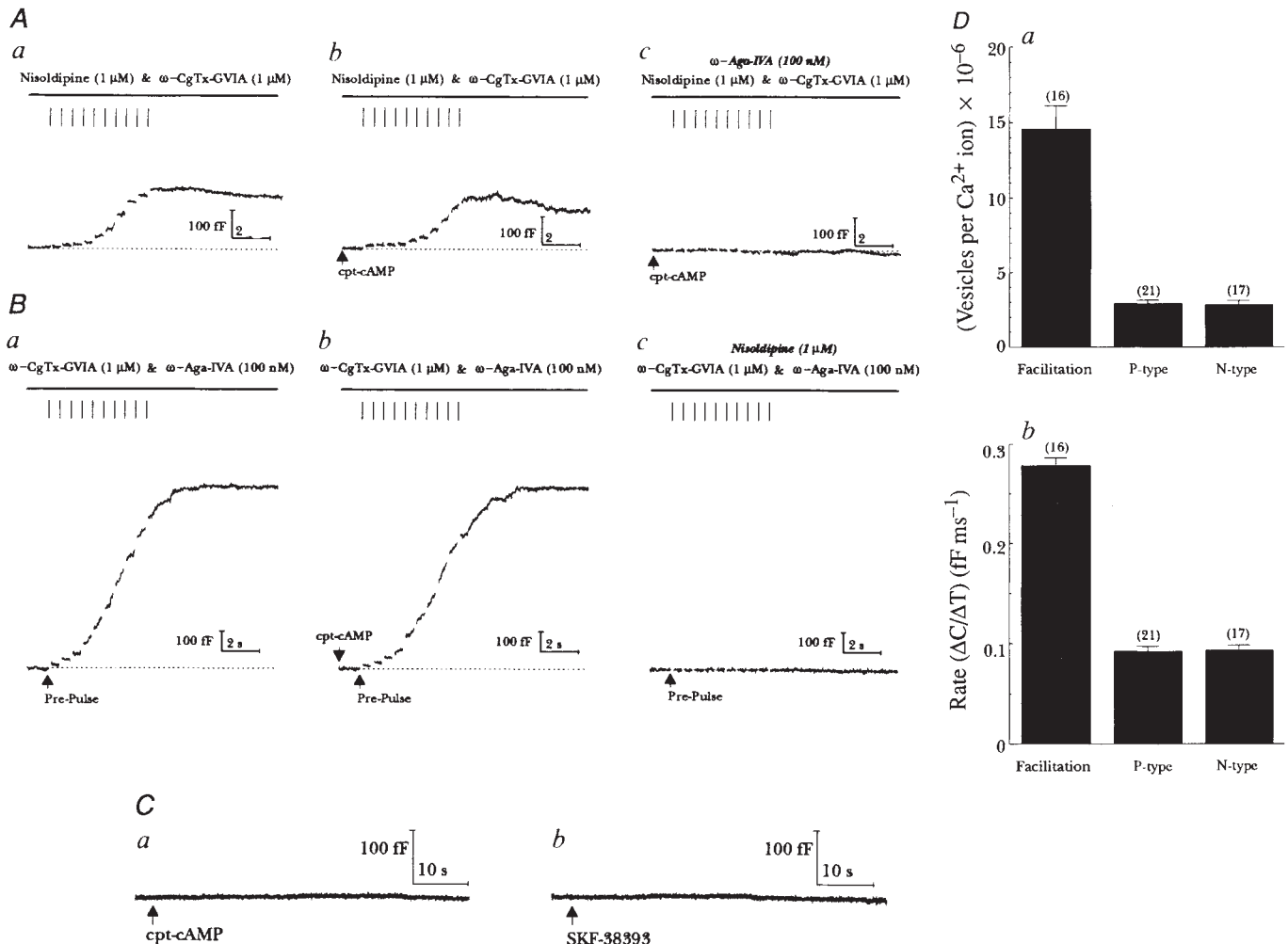


FIG. 3 Elevated $[cAMP]_i$ induces release exclusively as a result of recruiting facilitation Ca^{2+} currents. *A*, *a* and *b*, cAMP did not enhance secretion mediated by P-type Ca^{2+} channels. Capacitance changes were elicited by a train of depolarizations similar to that described for Fig. 1B. Nisoldipine and ω -CgTx suppressed facilitation and N-type currents. Traces were obtained before (*A*, *a*) and after (*A*, *b*) adding the membrane-permeant cAMP analogue 8-cpt-cAMP (100 μ M). At the end of this experiment, addition of ω -AGA-IVA completely suppressed secretion (*A*, *c*). *B*, *a* and *b*, Pre-pulses were used to recruit facilitation currents and elicit secretion. 8-cpt-cAMP and pre-pulses were not additive (*b*), suggesting that they were both working exclusively by recruiting facilitation currents (*a*). Nisoldipine (1 μ M) suppressed the secretion (*c*). Capacitance changes were again elicited by a train of 10 depolarizations. Each depolarization was preceded by a pre-pulse to +120 mV that lasted 50 ms. ω -CgTx and ω -Aga-IVA were used to suppress P- and N-type currents. *C*, *a* and *b*, $[cAMP]_i$ did not induce Ca^{2+} channel-independent secretion in chromaffin cells. Secretion was monitored as 8-cpt-cAMP (100 μ M; *a*) or a D_1 agonist (10 μ M; *b*) were added; no depolariza-

tions were applied. Note the scale bars for each panel are different. *D*, *a* Histogram of Ca^{2+} entry plotted as a function of the number of vesicles secreted, for each channel. Under our conditions, five times more Ca^{2+} ($\sim 3.5 \times 10^5$ Ca^{2+} ions) was required to release a single vesicle when P- or N-type Ca^{2+} channels were activated than when facilitation Ca^{2+} channels ($\sim 6.6 \times 10^4$) were involved. To calculate the number of vesicles released, the total capacitance change elicited by a train of depolarizations was divided by the average capacitance change produced by a single vesicle (2 fF)^{10,22}. Ca^{2+} -entry was obtained by integration of the area in the current records. *D*, *b*, Maximal change in capacitance as a function of time ($\Delta C/\Delta T$) for each Ca^{2+} current component indicates a ~ 3 -fold increase when facilitation channels were activated in comparison to P- or N-type channels. Maximal rates of secretion always occurred during the depolarizations. Note these results make no assumptions about the relation between Ca^{2+} influx and release. If release is related to Ca^{2+} influx in a nonlinear fashion, normalizing the data in Fig. 3 by this relation would tend to make the difference between channels even more dramatic.

P- or N-type channels trigger release less efficiently than facilitation Ca^{2+} channels; about five times the Ca^{2+} influx per vesicle is required (Fig. 3D, a). Figure 3D, b plots the rate of capacitance change as a function of time ($\Delta C/\Delta T$); although N- and P-type Ca^{2+} currents were larger in comparison to facilitation Ca^{2+} currents, the rate of capacitance change was smaller. In addition, P- or N-type Ca^{2+} channels showed a markedly increased latency for release when compared to facilitation channels.

From these results, one possibility is that the Ca^{2+} channels in chromaffin cells are not randomly distributed; facilitation Ca^{2+} channels may be closer to the docking and release sites than either of the other two channels. Indeed, a physical association of Ca^{2+} channels with components of the exocytotic apparatus has already been proposed^{2,15,16}. If all three channel types are randomly distributed, they might support secretion equally efficiently. The fact that this is not the case suggests that the precise localization of Ca^{2+} entry may be as important as its magnitude in determining its role in secretion. Nonetheless, physical association between Ca^{2+} channels and release sites alone cannot explain all our results. A remote priming site¹⁷ or different pools of vesicles in various stages of availability¹⁴ could account for the delay observed between depolarizing the cell and the onset of secretion. The large secretion observed in the 500 ms between depolarizations cannot be explained exclusively by a sustained increase in $[\text{Ca}^{2+}]$ (ref. 2), which will be similar for all three types of channels¹⁸; a maintained elevation of $[\text{Ca}^{2+}]$; and a local Ca^{2+} -dependent priming site together would better explain the observation. Other explanations for the difference in efficiency between distinct Ca^{2+} channel types and secretion are possible. Our results suggest that different Ca^{2+} channels in chromaffin cells regulate secretion depending on the level of electrical activity. Under conditions when few action potentials are generated, only N- and P-type channels would be expected to activate. Because channels are inefficiently coupled to exocytosis, little secretion occurs. Conditions of elevated electrical activity recruit

facilitation. Because facilitation Ca^{2+} channels are coupled more efficiently to secretion, catecholamine release can increase dramatically. The novel gating properties of facilitation Ca^{2+} channels, that is, recruitment by trains of depolarizations or by agents that increase $[\text{cAMP}]$, and their possible proximity to release sites suggests that these channels may be responsible for the dramatically augmented catecholamine secretion from the adrenal medulla that occurs during conditions of stress or danger. As most neurons and other secretory cells have more than one type of Ca^{2+} channel, the concept of differential coupling of specific channels to different rates of transmitter or hormone release may be generally applicable¹⁹. □

Received 9 July; accepted 19 November 1993.

- Artalejo, C. R., Mogul, D. J., Perlman, R. L. & Fox, A. P. *J. Physiol., Lond.* **444**, 213–240 (1991).
- Augustine, G. M. & Neher, E. *J. Physiol., Lond.* **450**, 247–271 (1992).
- Burgoyne, R. D. in *Neuromethods 20: Intracellular Messengers* (eds Boulton, A., Baker, G. & Taylor, C.) 433–469 (1992).
- Hans, M., Illes, P. & Takeda, K. *Neurosci. Lett.* **114**, 63–68 (1990).
- Artalejo, C. R., Perlman, R. L. & Fox, A. P. *Neuron* **8**, 85–95 (1992).
- Mintz, I. M. et al. *Nature* **355**, 827–829 (1992).
- Artalejo, C. R., Dahmer, M. K., Perlman, R. L. & Fox, A. P. *J. Physiol., Lond.* **432**, 681–707 (1991).
- Artalejo, C. R., Rossie, S., Perlman, R. L. & Fox, A. P. *Nature* **358**, 63–66 (1992).
- Artalejo, C. R., Ariano, M., Perlman, R. L. & Fox, A. P. *Nature* **348**, 239–242 (1990).
- Neher, E. & Marty, A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6712–6716 (1982).
- Mintz, I. M., Adams, M. E. & Bean, B. P. *Neuron* **9**, 85–95 (1992).
- Amalia, C., Ashcroft, F. M. & Rorsman, P. *Nature* **363**, 356–358 (1993).
- Morgan, A., Wilkinson, M. & Burgoyne, R. D. *EMBO J.* **12**, 3747–3751 (1993).
- Neher, E. & Zucker, R. S. *Neuron* **10**, 21–30 (1993).
- Pumplin, D. W., Reese, T. S. & Llinas, R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7210–7213 (1981).
- Bennett, M. K., Calakos, N. & Scheller, R. H. *Science* **257**, 255–259 (1992).
- Nowicky, M. C. *Biophys. J.* **64**, A230 (1992).
- Nowicky, M. C. & Pinter, M. J. *Biophys. J.* **64**, 77–91 (1993).
- Lemos, J. M. & Nowicky, M. C. *Neuron* **2**, 1419–1426 (1989).
- Joshi, C. & Fernandez, J. M. *Biophys. J.* **53**, 885–892 (1988).
- Fidler, N. H. & Fernandez, J. M. *Biophys. J.* **56**, 1153–1162 (1989).
- Coupland, R. E. *Nature* **217**, 384–388 (1986).

ACKNOWLEDGEMENTS. We thank C. Palfrey for suggestions and for critically reading the manuscript, and E. Zwerger for the preparation of cells. This work was supported by the Alfred P. Sloan Foundation (C.R.A.), the NIH (M.E.A. and A.P.F.) and the Brain Research Foundation (A.P.F.).

Components of wingless signalling in *Drosophila*

Esther Siegfried, Elizabeth L. Wilder
& Norbert Perrimon*

Department of Genetics, * Howard Hughes Medical Institute, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

THE determination of specific cell fates and polarity within each segmental unit of the *Drosophila* embryo involves the products of the segment polarity genes¹. One of these, *wingless* (*wg*), encodes a secreted protein^{2,3} that is homologous to the mammalian proto-oncogene Wnt-1 (refs 4, 5). In the embryonic epidermis, *wg* is expressed in a single row of cells within each segmental unit, although its activity is required for the correct patterning of most of the epidermis^{4,6}. Initially Wg signals to adjacent posterior cells, maintaining *engrailed* (*en*) expression^{7,8}. Later during embryogenesis, *wg* specifies the differentiation of naked cuticle⁹. Wg signalling functions by inactivating or antagonizing the activity of *zeste-white 3* (*zw3*)¹⁰. We have investigated the requirement in the Wg signal transduction pathway for the three genes *armadillo* (*arm*)^{11,12}, *dishevelled* (*dsh*) and *porcupine* (*porc*)¹³, all of which have embryonic mutant phenotypes similar to *wg*. Our results indicate that *dsh* and *porc* act upstream of *zw3*, and *arm* acts downstream of *zw3*.

Embryos that lack *wg* activity (referred to as *wg* embryos) are completely covered with denticles on the ventral cuticle, in

contrast to wild-type embryos, in which the ventral cuticle is an alternating pattern of naked cuticle and denticle belts (Fig. 1a)¹⁴. The wild-type cuticular pattern depends on the expression of both *wg* and *en*. These genes are expressed in adjacent stripes within each segmental unit and their expression is mutually dependent; in *wg* embryos the striped pattern of *en* in the epidermis decays and in *en* embryos the epidermal pattern of Wg decays^{7,8,15}. Although Wg signalling is essential⁶, its mechanism is unclear. *arm* is a member of a multigene family that includes the vertebrate plakoglobin and β -catenin proteins, involved in cell-cell adhesion^{16–18}. *dsh* encodes a novel protein of unknown biochemical function¹⁹ and *porc* has not been characterized. Genetic epistasis can be used to understand the function of these genes in Wg signalling. However, analysis of embryos doubly mutant for *wg* and any single gene are uninformative because of the similarity of their phenotypes. Double-mutant combinations with *zw3* are revealing, because loss of *zw3* activity results in an embryonic phenotype opposite to that of *wg* embryos; *zw3* embryos lack denticles on the ventral cuticle (Fig. 1e) and each of the epidermal *en* stripes is expanded (Fig. 2e)²⁰. *zw3*, another component of Wg signalling, encodes the *Drosophila* homologue of the mammalian serine/threonine protein kinase glycogen synthase kinase 3 (refs 21–23).

Embryos that are doubly mutant for *zw3* and *dsh* resemble *zw3* embryos; they lack denticles on the ventral cuticle (Fig. 1f), as observed in *zw3* embryos (Fig. 1e) and in contrast to *dsh* embryos (Fig. 1b). In *zw3 dsh* embryos, the pattern of *en* expression is expanded in a manner identical that seen in *zw3* embryos (compare Fig. 2e with f). After *en* expression expands, ectopic stripes of Wg appear in *zw3* and *zw3 dsh* embryos (compare Fig. 3f with g)^{10,20}. This is in contrast to *dsh* embryos, where Wg is

not detected in the epidermis at this stage of embryogenesis (Fig. 3b)¹⁹. The similarity of the phenotypes of *zw3* and *zw3 dsh* embryos suggests that *dsh* functions upstream of *zw3* in Wg signalling.

Conversely, *arm zw3* double-mutant embryos exhibit the *arm* cuticle phenotype, the ventral cuticle is covered with denticles (compare Fig. 1c with g). But in the double-mutant embryos, En and Wg are detected in the epidermis at a stage where they are no longer expressed in *arm* embryos (compare Fig. 2c with g, and Fig. 3c with h)^{11,12}. The stripes of En staining in *arm zw3* embryos are not as broad as in *zw3* embryos (Fig. 2e) and they eventually fade (data not shown). The cuticular phenotype of *arm zw3* embryos suggests that *arm* function is downstream of *zw3* in Wg signalling. The epistasis is not entirely clear because the patterns of En and Wg staining in the doubly mutant embryos do not resemble the patterns observed in *arm* embryos. The phenotype we observe in *arm zw3* embryos may reflect residual activity of the temperature-sensitive mutation, *arm*^{H8.6}, which we used. Possibly different levels of Arm activity are needed for the expression of En versus the specification of naked cuticle. In an independent study, using *arm* alleles that are not temperature sensitive, *arm zw3* embryos were found to resemble *arm* embryos for the cuticular phenotype, as well as the pattern of En expression²⁴. These results suggest that *zw3* function is mediated through Arm.

We have also examined the relationship between *zw3* and *arm* in another fashion. *wg* regulates *arm* post-transcriptionally, resulting in the appearance of alternating stripes and interstripes of Arm protein in the epidermis of wild-type embryos (Fig. 4Aa)²⁵. To determine whether the *wg*-dependent accumulation of Arm is regulated by *zw3* activity, we examined the distribution of Arm in *zw3* embryos. The pattern of stripes and interstripes of Arm is lost in *zw3* embryos, and the protein is uniformly distributed at high levels (Fig. 4Ab). The accumulation of Arm into stripes, relative to interstripes, reflects both increased levels and distinct subcellular localization of the protein²⁴. Furthermore these authors demonstrate that both the levels as well as the subcellular localization of Arm protein are altered in *zw3* embryos.

Finally we examined embryos lacking both *zw3* and *porc* activity. *zw3 porc* embryos exhibit a cuticular phenotype similar to *zw3* embryos: the ventral cuticle is devoid of denticles (compare Fig. 1e with h). The pattern of En and Wg stripes are identical in *zw3 porc* embryos and *zw3* embryos (Figs 2e, h and 3f, i). These results suggest that *porc* functions upstream of *zw3* in the transduction of the Wg signal.

Although embryos that are mutant for *dsh*, *arm* or *porc* exhibit similar phenotypes, they differ in the pattern of Wg staining. Epidermal Wg staining can be detected in *porc* embryos at a developmental stage where it has faded in *dsh* and *arm* embryos.

FIG. 1 The cuticle phenotype of wild-type (a), *dsh*^{V26} (b), *arm*^{H8.6} (c), *porc*^{PB16} (d), *zw3*^{M11-1} (e), *zw3*^{M11-1} *dsh*^{V26} (f), *arm*^{H8.6} *zw3*^{M11-1} (g) and *zw3*^{M11-1} *porc*^{PB16} (h) mutant embryos. All the mutant embryos shown are derived from homozygous mutant germ lines and have received neither maternal or zygotic wild-type product. All experiments were at 25 °C. In general the phenotype of *arm zw3* embryos is not as severe as that of *arm* embryos. Embryos are shown anterior up and ventral to the left.

METHODS. Cuticles were prepared and mounted in Hoyer's medium²⁸ mixed 1:1 with lactic acid, viewed and photographed under dark field. Mutant embryos are derived from females with mosaic germ lines generated by the 'FLP-DFS' technique²⁹. Females with germ-line clones were of the genotypes *zw3*^{M11-1} *FRT*¹⁰¹/*ovo*^{D1} *FRT*¹⁰¹; *FLP*^{F38}/+, *dsh*^{V26} *FRT*¹⁰¹/*ovo*^{D1} *FRT*¹⁰¹; *FLP*^{F38}/+, *arm*^{H8.6} *FRT*¹⁰¹/*ovo*^{D1} *FRT*¹⁰¹; *FLP*^{F38}/+, *porc*^{PB16} *FRT*⁹⁻²/*ovo*^{D2} *FRT*⁹⁻²; *FLP*^{F38}/+, *zw3*^{M11-1} *dsh*^{V26} *FRT*¹⁰¹/*ovo*^{D1} *FRT*¹⁰¹; *FLP*^{F38}/+, *arm*^{H8.6} *zw3*^{M11-1} *FRT*¹⁰¹/*ovo*^{D1} *FRT*¹⁰¹; *FLP*^{F38}/+, *zw3*^{M11-1} *porc*^{PB16} *FRT*⁹⁻²/*ovo*^{D2} *FRT*⁹⁻²; *FLP*^{F38}/+. Descriptions of *ovo*^{D1} *FRT*¹⁰¹, *FLP*^{F38} and the production of germ line clones, with the exception of *porc* and *zw3 porc*, have been given previously²⁹. *FRT*⁹⁻² chromosome has an insertion of an FRT element at position 18E on the polytene chromosome map (T. B. Chou, E. Noll and N.P., unpublished data). *ovo*^{D2} *FRT*⁹⁻² is a recombinant chromosome between the female sterile mutation *ovo*^{D2} (ref. 30) and *FRT*⁹⁻², provided by E. Wieschaus. To generate germ-line clones, larvae of the genotype *porc*^{PB16} *FRT*⁹⁻²/*ovo*^{D2} *FRT*⁹⁻²; *FLP*^{F38}/+ and *zw3*^{M11-1} *porc*^{PB16} *FRT*⁹⁻²/*ovo*^{D2} *FRT*⁹⁻²; *FLP*^{F38}/+ were heat-shocked at 37 °C three times during larval development, for 2 h each time. *dsh*, *arm* and *porc* are completely paternally rescued, such that germ-line clone-derived embryos which receive one wild-type copy of the gene paternally are wild-type. *zw3* is only partially paternally rescued¹⁰.

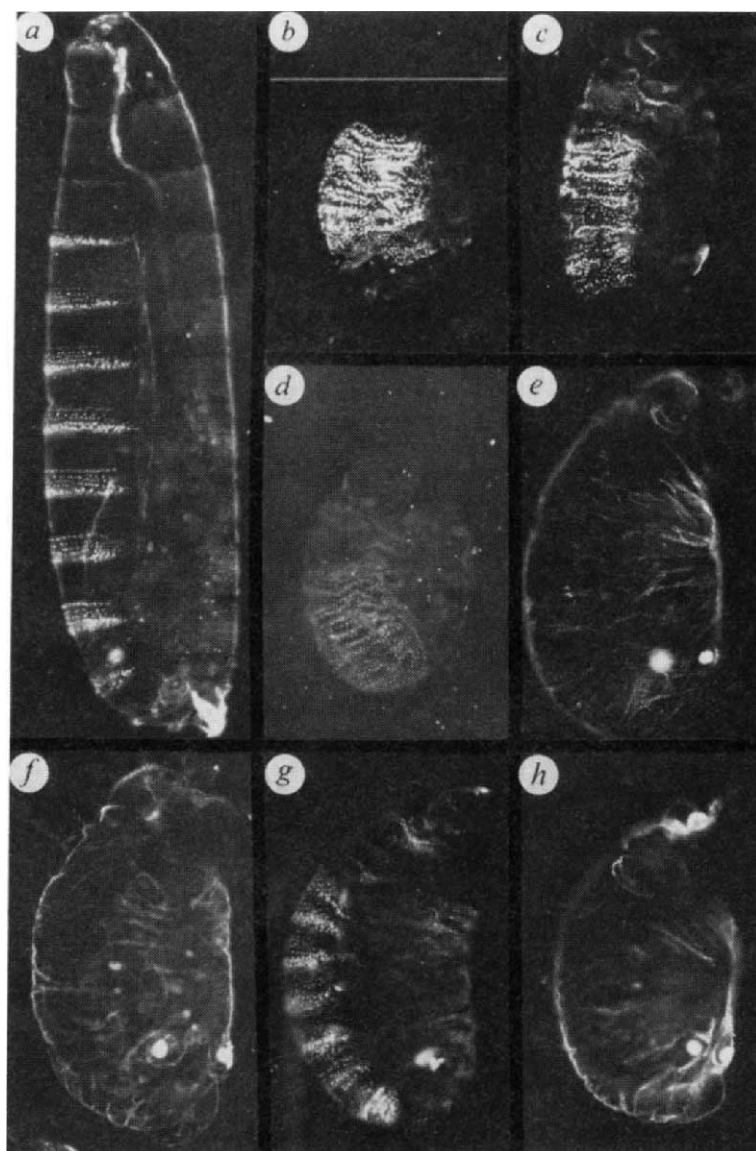


FIG. 2 Engrailed immuno-staining in stage 10 wild-type (a), *dsh*^{V26} (b), *arm*^{H8.6} (c), *porc*^{PB16} (d), *zw3*^{M11-1} (e), *zw3*^{M11-1} *dsh*^{V26} (f), *arm*^{H8.6} *zw3*^{M11-1} (g) and *zw3*^{M11-1} *porc*^{PB16} (h) mutant embryos. En expression has disappeared from the epidermis of *dsh*, *arm* and *porc* embryos but can still be detected in the nervous system. In contrast, En can be detected in the epidermis of *zw3* *dsh* and *zw3* *porc* embryos as observed in *zw3* embryos. In *arm* *zw3* embryos the En stripe is not as wide as in *zw3* embryos and does eventually fade (data not shown). Embryos are shown anterior up and ventral to the left. All experiments were at 25 °C.

METHODS. Females of the genotypes described in Fig. 1 were mated to FM7, *ftz-lacZ*/Y males³¹; this FM7 chromosome carries a transposon that contains the *fushi tarazu* promoter fused to the *Escherichia coli* LacZ gene. This allowed us to distinguish the genotypes of embryos derived from germ-line clones. Paternally rescued embryos that have received one wild-type copy of the gene from the father stained for β -galactosidase. Null embryos that received neither maternal or paternal wild-type gene product do not stain for β -galactosidase. The embryos were collected, dechorionated in 50% bleach and fixed for 10–15 min in 4% formaldehyde in PBS. Embryos were first stained for X-gal to detect the paternally rescued embryos and then immunostained using a mouse monoclonal antibody directed against En³², diluted 1:1, as previously described¹⁰. Embryos were mounted and viewed in methyl salicylate. Only the null mutant embryos are shown. All staging of embryos is according to previous descriptions³³.

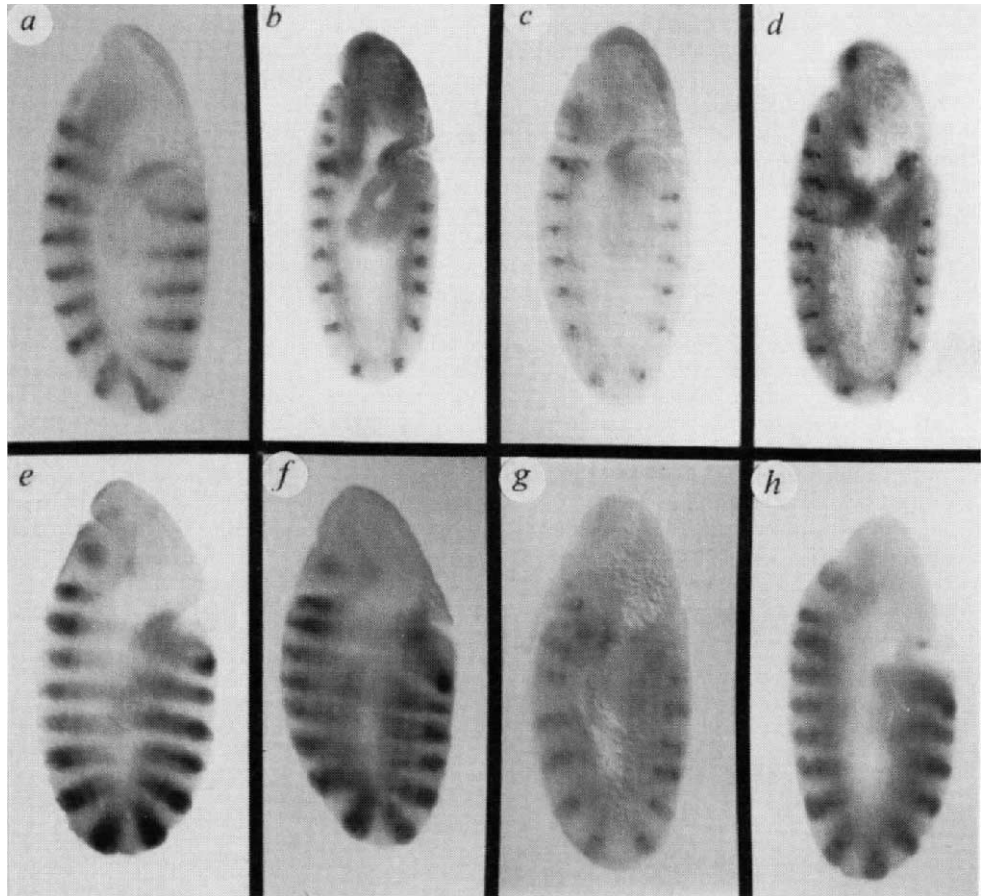
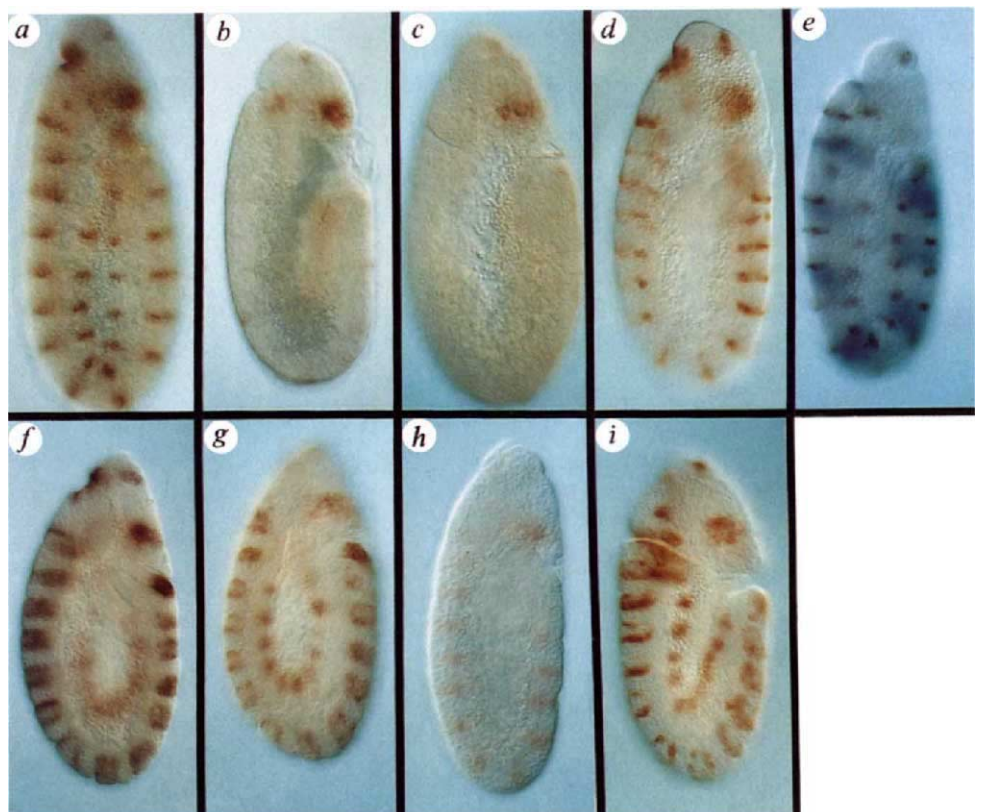


FIG. 3 Wingless immunostaining in stage 11 wild-type (a), *dsh*^{V26} (b), *arm*^{H8.6} (c), *porc*^{PB16} (d), *porc*^{PB16} rescued (e), *zw3*^{M11-1} (f), *zw3*^{M11-1} *dsh*^{V26} (g), *arm*^{H8.6} *zw3*^{M11-1} (h) and *zw3*^{M11-1} *porc*^{PB16} (i) mutant embryos. Embryos are shown anterior up and ventral to the left. All experiments were at 25 °C. Wg staining decays in *dsh* and *arm* embryos, but can still be detected in *porc* embryos. In *arm* *zw3* embryos only the wild-type stripe of Wg can be detected. In *porc* embryos, as well as *porc*-rescued embryos, Wg staining is discrete, the protein appears to be retained by the cells that synthesize it. *zw3* *dsh* and *zw3* *porc* embryos show an ectopic stripe of Wg, as observed in *zw3* embryos.

METHODS. Females of the genotypes described in Fig. 1 were mated to FM7, *ftz-lacZ*/Y males. The embryos were collected, dechorionated in 50% Chlorox and fixed for 35 min in 4% formaldehyde in PBS. Polyclonal serum directed against Wg² was used at a 1:500 dilution and immunostaining was detected with peroxidase-conjugated goat anti-rabbit secondary antibody (diluted 1:500). Rescued and null mutant embryos were distinguished by staining using mouse monoclonal antibody against β -galactosidase, 1:1,000 dilution, and detected with alkaline phosphatase conjugated goat anti-mouse secondary antibody (diluted 1:500). Embryos were mounted and viewed in 70% glycerol. Wg immunostaining is brown and β -galactosidase immunostaining is blue.



In addition, in *porc* embryos Wg protein appears confined to the *wg*-expressing cells, leading to the suggestion that in *porc* embryos Wg is retained by the cells that synthesize it (Fig. 3d)²⁶. The *porc* embryonic segment polarity phenotype, which results from the lack of both maternal and zygotic product, can be completely rescued to adulthood by one wild-type copy of the

zygotic gene¹³. We were therefore surprised to observe that *porc*-rescued embryos also exhibit confined Wg staining, indicative of protein retention (Fig. 3e). Later in development (stage 13), Wg does appear to be secreted in *porc*-rescued embryos (data not shown). These results suggest that early zygotic expression of *porc* enables low levels of Wg to be secreted or that it allows Wg to reach the cell surface, where it is sufficient to signal to adjacent cells and consequently maintain En expression.

We conclude that both *porc* and *dsh* activities are required upstream of *zw3* in Wg signalling. Our data are consistent with a proposed role for Porc in the secretion of Wg protein and a role for Dsh in the reception or transduction of the Wg signal. *arm* activity is required downstream of *zw3* and our results, as well as others, suggest that *zw3* activity is mediated through a redistribution of Arm protein. We propose that these gene products interact in a linear pathway based on the epistatic relationships we have determined. But it is possible that these genes function in parallel pathways. This seems unlikely in the case of *zw3* and *arm* because Arm protein distribution is altered in *zw3* embryos. Our conclusions are also consistent with recent work that indicates that both *dsh* and *arm*, but not *porc*, are required for the gain-of-function phenotype generated by ubiquitous embryonic expression of *wg* (ref. 27).

We propose the following model for Dsh, Arm and Porc function in the Wg signal transduction pathway (Fig. 4B). Porc is necessary for the correct distribution of Wg protein. Wg protein is apparently secreted by cells that synthesize it and received by neighbouring cells^{2,3}, where a signal transduction cascade is initiated. Once received by cells, the Wg signal is transduced through Dsh, which results in the inactivation of Zw3 kinase, which in turn modulates the level of active Arm protein. The correct epidermal En expression, as well as the specification of naked cuticle, is achieved by the modulation of Arm. Zw3 kinase may directly phosphorylate Arm, thereby reducing the levels of active protein, whereas Wg signalling can abrogate this effect by inactivating Zw3 kinase, resulting in high levels of active Arm. This hypothesis is consistent with the observation that Arm is a phosphoprotein and that Zw3 kinase is required for its phosphorylation (M. Peifer, personal communication).

But how might Arm, a molecule involved in cell adhesion, regulate the expression of *en*? Simplistically, high levels of active Arm protein may directly transduce the Wg signal from the membrane to the nucleus. Alternatively, Arm accumulation may be required to make specific intercellular contacts, and these contacts are a prerequisite for a second process that would regulate *en* expression and cell fate determination. □

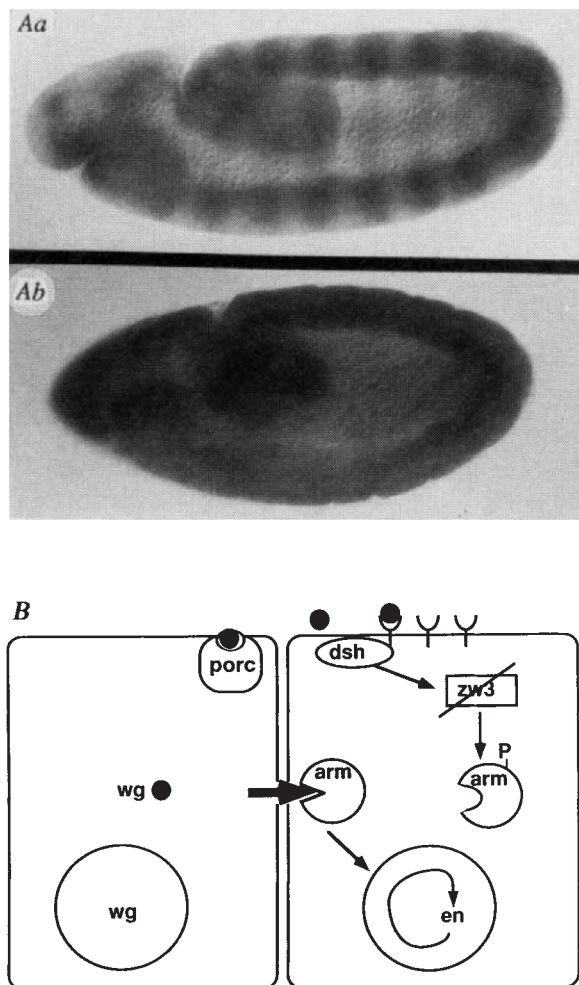


FIG. 4 A, Armadillo immunostaining in stage 10 wild-type (A, a) and *zw3* mutant (A, b) embryos. Embryos are shown anterior to the left and dorsal up. In wild-type embryos a low level of Arm is detected throughout the epidermis, in addition to stripes of high levels of staining positioned over the parasegmental border²⁵. In *zw3* embryos a high level of Arm staining is detected throughout the epidermis. B, Model for the mechanism of Wg signalling. The maintenance of *en* expression requires input from Wg (shown as black dots). The distribution of Wg protein is regulated by Porc. Dsh is required for the transduction of the Wg signal in adjacent cells, perhaps in association with the unidentified Wg receptor. Transduction of the Wg signal inactivates or antagonizes Zw3 kinase, which functions to modulate the levels of active Arm. Zw3-dependent phosphorylation of Arm results in low levels of active protein, whereas Wg signalling can revert this effect and result in accumulation of active Arm. Arm either directly, or through the formation of a second signalling complex (large arrow), results in the maintenance of En expression.

METHODS. Embryos derived from *zw3* mutant germ lines were generated as described for Fig. 2. Embryos were collected, decolorized and fixed as described for Fig. 3. Rabbit polyclonal antibody against Arm²⁵, kindly provided by M. Peifer, was used at a 1:500 dilution. Embryos were then incubated with a 1:500 dilution of biotinylated goat anti-rabbit secondary, followed by incubation with Vectastain. Embryos were mounted and viewed in methyl salicylate.

Received 7 June; accepted 19 October 1993.

- Ingham, P. W. *Curr. Opin. Genet. Dev.* **1**, 261–267 (1991).
- Van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. *A. Cell* **59**, 739–749 (1989).
- Gonzalez, F., Swales, L., Bejsovec, A., Skaer, H. & Martinez-Arias, A. *Mech. Dev.* **35**, 43–54 (1991).
- Baker, N. E. *EMBO J.* **6**, 1765–1773 (1987).
- Rijsewijk, F. et al. *Cell* **50**, 649–657 (1987).
- Peifer, M. & Bejsovec, A. *Trends Genet.* **8**, 243–249 (1992).
- Martinez-Arias, A., Baker, N. & Ingham, P. W. *Development* **103**, 153–170 (1988).
- DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassisi, J. A. & O'Farrell, P. H. *Nature* **332**, 604–609 (1988).
- Bejsovec, A. & Martinez-Arias, A. *Development* **113**, 471–485 (1991).
- Siegrfried, E., Chou, T. B. & Perrimon, N. *Cell* **71**, 1167–1179 (1992).
- Klingensmith, J., Noll, E. & Perrimon, N. *Dev. Biol.* **134**, 130–145 (1989).
- Peifer, M., Rauskolb, C., Williams, M., Riggelman, B. & Wieschaus, E. *Development* **111**, 1029–1043 (1991).
- Perrimon, N., Engstrom, L. & Mahowald, A. P. *Genetics* **121**, 333–352 (1989).
- Nusslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795–801 (1980).
- Heemskerk, J., Di Nardo, S., Kostriken, R. & O'Farrell, P. *Nature* **352**, 404–410 (1991).
- Peifer, M. & Wieschaus, E. *Cell* **63**, 1167–1178 (1990).
- McCrea, F. D., Turck, C. W. & Gumbiner, B. *Science* **254**, 1359–1361 (1991).
- Franke, W. W. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4027–4031 (1989).
- Klingensmith, J., Nusse, R. & Perrimon, N. *Genes Dev.* (in the press).
- Perrimon, N. & Smouse, D. *Dev. Biol.* **135**, 287–305 (1989).
- Siegrfried, E., Perkins, L. A., Capaci, T. M. & Perrimon, N. *Nature* **345**, 825–829 (1990).
- Bourouis, M. et al. *EMBO J.* **9**, 2877–2884 (1990).
- Woodgett, J. R. *Trends biochem. Sci.* **16**, 177–181 (1991).
- Peifer, M., Sweeton, D., Casey, M. & Wieschaus, E. *Development* (in the press).
- Riggelman, B., Schedl, P. & Wieschaus, E. *Cell* **63**, 549–560 (1990).

26. Van den Heuvel, M., Harryman-Samos, C., Klingensmith, J. & Nusse, R. *EMBO J.* (in the press).
27. Noordermeer, J., Klingensmith, J., Perrimon, N. & Nusse, R. *Nature* **367**, 80–83 (1994).
28. Van der Meer, J. *Dros. Inf. Serv.* **52**, 160 (1977).
29. Chou, T. B. & Perrimon, N. *Genetics* **131**, 643–653 (1992).
30. Perrimon, N. *Genet.* **108**, 927–939 (1984).
31. Kania, M. A., Bonner, A. S., Duffy, J. B. & Gergen, J. P. *Genes Dev.* **4**, 1701–1713 (1990).
32. Patel, N. et al. *Cell* **58**, 955–968 (1989).
33. Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, New York, Berlin, 1985).

ACKNOWLEDGEMENTS. We thank T. B. Chou, P. Gergen, N. Patel, M. Peifer, M. van den Heuvel and E. Wieschaus for reagents, J. Klingensmith, J. Noordermeer, R. Nusse, M. Peifer and E. Wieschaus for communication of unpublished results, and A. Lassar, J. Noordermeer, R. Nusse, L. Perkins, R. Riddle, G. Thomas and K. Yoffe for comments on the manuscript. This work was supported by a Charles King Trust/Medical Foundation fellowship to E.S., an NIH fellowship to E.L.W. and The March of Dimes Foundation; N.P. is an Investigator of the Howard Hughes Medical Institute.

dishevelled and armadillo act in the Wingless signalling pathway in *Drosophila*

Jasprien Noordermeer*, John Klingensmith*†, Norbert Perrimon† & Roel Nusse*†

* Howard Hughes Medical Institute and Department of Developmental Biology, Stanford University, California 94305-5428, USA

† Howard Hughes Medical Institute and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

THE *Wnt* genes encode conserved secreted proteins that play a role in normal development and tumorigenesis^{1,2}. Little is known about the signal transduction pathways of *Wnt* gene products. One of the best characterized *Wnt* family members is the *Drosophila* segment polarity gene *wingless*^{3–6}. We have investigated whether segment polarity genes with a *wingless*-like phenotype mediate the *wingless* signal. We used a *wingless* transgene controlled by a heat-shock promoter for genetic epistasis experiments. We show that *wingless* acts through *dishevelled* and *armadillo* to affect the expression of the homeobox gene *engrailed* and cuticle differentiation.

During germ-band extension, *wingless* (*wg*) is expressed in stripes flanking a parasegmental border³. The Wg protein is secreted^{7,8} and taken up by adjacent cells for maintenance of expression of the homeobox gene *engrailed* (*en*)^{9–12}. Later, *wg* is necessary for the generation of smooth cuticle¹³. No naked cuticle is formed in the absence of *wg*; instead the ventral cuticle consists of a lawn of denticles.

Genes that mediate the *wg* signal are probably among the segment polarity mutants with a *wg*-like phenotype. Absence of functional products of the genes *dishevelled* (*dsh*), *armadillo* (*arm*), *porcupine* (*porc*) and *hedgehog* (*hh*) lead to cuticle defects similar to that of *wg* mutant embryos and also to a loss of *en* expression^{9,14–16}. It has been difficult to order these mutations in a genetic pathway because their phenotypes are nearly indistinguishable and because expression of *wg* depends on *en* expression. But the order of action of two genes in a common pathway can be found by examining epistasis between a dominant gain-of-function mutant for one gene in combination with a loss-of-function mutant for the other. Dominant *wg* alleles have not been identified, but we previously established a transgenic *Drosophila* strain that expresses *wg* under the control of a heat-shock promoter¹⁷. Heat-shock during particular stages of development causes expansion of the *En* expression domain and generation of a completely naked ventral cuticle¹⁷ (Figs 1b and 2b). We used the HS-*wg* allele to determine which of several segment polarity genes with a *wg*-like phenotype are required for *wg* action.

We first investigated whether the effects of HS-*wg* on expansion of *En* expression require endogenous *wg* and *en* genes. The *en*^{CS1} allele produces a cytoplasmic, non-functional protein, in contrast to the nuclear localization of the *En* protein in wild type. In *en*^{CS1} embryos, *En* expression is not maintained¹¹. The heat-shocked double mutant *en*^{CS1};HS-*wg* embryos can be distinguished from the single mutants by their unique pattern of *En* protein expression, which is ectopically induced at early stage 10 (ref. 18) as in HS-*wg* embryos, but localized in the cytoplasm as in *en*^{CS1} (Fig. 1i, j). Normal or ectopic *En* domains are not maintained: at stage 11 most *En* protein has disappeared from the ectoderm (data not shown).

In HS-*wg* embryos, an ectopic domain of Wg protein made from the normal *wg* gene is formed just posterior to the expanded *En* domain¹⁷. In the heat-shocked double mutant *wg*;HS-*wg* embryos, the ectopic Wg domain does not appear, yet the expanded domain of *En* expression is induced (compare Fig. 1a with h) and maintained (data not shown)²⁰. Thus neither functional *En* nor Wg protein made from the endogenous *wg* gene are required to mediate the effects of HS-*wg* on induction of ectopic *En*.

To examine whether genes with mutant phenotypes similar to *wg* (*dsh*, *arm*, *porc* and *hh*)^{14–16,19} are required for the effects of HS-*wg* on *En* expression, we made double mutants between HS-*wg* and loss-of-function mutations in these genes. As *Dsh*, *Arm* and *Porc* products are provided both maternally and zygotically, germ-line mosaic females were derived to remove completely the gene product from the developing embryo. In all four single mutants, *En* expression decays as in *wg* embryos (Fig. 1c), but there are some differences in pattern and timing of *En* decay^{9,14}. The *En* patterns in the double mutant embryos are shown in Fig. 1d–g. In *dsh*;HS-*wg* embryos (Fig. 1d), the *En* expression pattern is very similar to *dsh* and *wg* embryos, although dorsally the *En* protein disappears from the ectoderm slightly later in development than in *dsh* embryos. Likewise, *arm*;HS-*wg* embryos (Fig. 1e) show a pattern of *En* distribution similar to *arm* embryos. In contrast, *porc*;HS-*wg* (Fig. 1f) and *hh*;HS-*wg* embryos (Fig. 1g) show ectopic *En* expression as seen in HS-*wg* embryos. Thus *dsh* and *arm* are required for induction of ectopic *En* in HS-*wg*, whereas *hh* and *porc* are not.

We then studied the effects of the absence of *dsh*, *arm*, *porc* or *hh* on cuticle pattern formation in HS-*wg*. The cuticle patterns of *dsh*;HS-*wg* (Fig. 2d) and *arm*;HS-*wg* (Fig. 2e) embryos are indistinguishable from the germ-line clone-derived *dsh* or *arm* embryos and from *wg* embryos (Fig. 2c) and develop a continuous lawn of denticles. In contrast, *porc*;HS-*wg* embryos show some restoration of the segmental denticle pattern (Fig. 2f). *hh*;HS-*wg* embryos (Fig. 2g) have a similar but not identical cuticle pattern to HS-*wg* (Fig. 2b)²⁰: naked throughout most of the ventral cuticle, with dispersed patches of non-polarized denticles mostly present at the lateral sides. Thus *dsh* and *arm*, but not *porc* and *hh*, are epistatic to and presumably downstream of HS-*wg*, not only in the effect on *En* expression but also in the generation of naked cuticle. The cuticle of *wg*;HS-*wg* double-mutant embryos is shown in Fig. 2h. After a 20-min heat shock during germ-band extension, an embryo of almost wild-type size is formed with partially restored head and tail structures (filzkörper) and a segmental pattern. This unexpected result suggests that differential levels of *wg* are not essential for at least some of its functions²⁰. It is possible that *wg* normally regulates pattern together with other determinants that function in a spatially restricted manner. For example, not all cells may be equally sensitive to *wg* activity, because of differences in concentrations of interpreting molecules.

In conclusion, *dsh* and *arm* are essential components of the *wg* signalling pathway and probably act downstream of *wg*. In contrast to the non-autonomous behaviour of *wg* in mutant cell clones, *dsh* and *arm* act autonomously, suggesting a role in reception of the *wg* signal^{16,21}. In addition, *wg* is required for post-transcriptional regulation of *arm*, resulting in an accumulation

† To whom correspondence should be addressed.

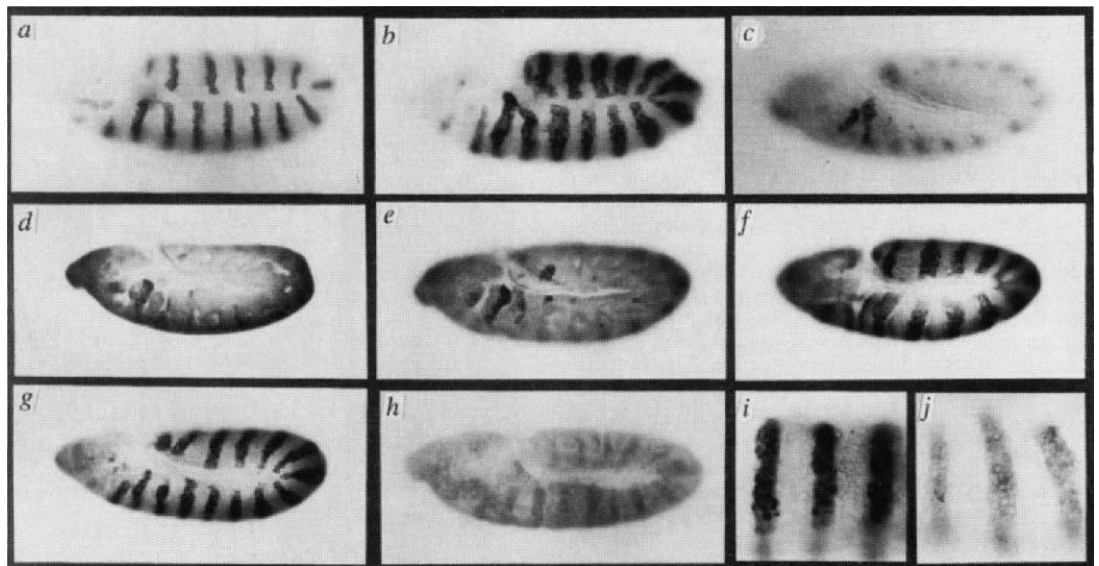
or redistribution of Arm protein in segmentally repeated stripes in the Wg domain²². The product of the *arm* gene is homologous to the vertebrate proteins plakoglobin and β -catenin²³, components of cell-cell junctions. Post-transcriptional modification of *arm* by *wg* requires *dsh*, suggesting that *dsh* functions upstream of *arm*²². The *dsh* gene encodes a novel protein that is highly conserved in evolution²¹.

porc, like *wg*, behaves non-autonomously in mitotic cell clones (J.A.K. and N.P., manuscript in preparation). In *porc* mutant embryos, the distribution of the Wg protein is altered in a way that suggests secretion of the protein is impaired²⁴. Possibly, *porc* is required for paracrine functions of *wg*; by expressing *wg* within the *en* cell (under the control of the ubiquitously acting heat-shock promoter), the requirements for *porc* can be bypassed,

resulting in induction of ectopic En. But the cuticle phenotype of *porc*;HS-*wg* is not like HS-*wg*, showing a slight restoration of pattern reminiscent of that of *wg*;HS-*wg* embryos, maybe indicating that endogenous *wg* function in these embryos is impaired. We found that Hh protein, a putative transmembrane protein²⁵, is not necessary for the effects of HS-*wg*. On the basis of genetic evidence, it has been suggested that *hh* is a signal from the *en* cells to maintain Wg expression²⁶. *hh* would therefore act upstream of *wg*, which is consistent with our findings.

wg and *dsh* act upstream of *zeste-white 3* (*zw3*), a serine/threonine kinase and repressor of *en*²⁷⁻²⁹, whereas *arm* is epistatic to *zw3* (ref. 30). Taken together with our results, a pathway can now be suggested in which *hh* acts upstream of *wg*, *dsh* downstream of *wg* and upstream of *zw3*, whereas *arm* and *en* act

FIG. 1 Effects of the lack of *dsh*, *arm*, *porc*, *hh*, endogenous *wg* or functional *en* on expansion of En in stage 10 (ref. 18) HS-*wg* embryos. Anterior is to the left, dorsal is up. *a-h*, Surface views of whole-mount embryos; *i, j*, higher magnification surface views of segments T₁ to T₃. All embryos were heat-shocked as described¹⁷. *a-c*, En expression pattern in wild-type (*a*), HS-*wg* (*b*) and *wg*^{cx4} embryos (*c*). *d-f*, Progeny of crosses between germ-line mosaic females of *dsh*⁴⁷⁷, *arm*^{25B} or *porc*^{PB16} with HS-*wg* males carrying an *FM7*, *ftz LacZ* balancer were stained for both En and β -galactosidase. In all cases



lack of β -galactosidase staining was used as a marker to identify the segment-polarity mutant embryos. *d*, Embryo from the *dsh* × *FM7*, *ftz LacZ*/Y;HS-*wg*/+ cross; all germ-line clone embryos without *dsh* function lose En expression. The same result is obtained in the cross *arm* × *FM7*, *ftz LacZ*/Y;HS-*wg*/+ (*e*). In the *porc* × *FM7*, *ftz LacZ*/Y;HS-*wg* cross, 50% of the germ-line clone embryos without *porc* product show induction of ectopic En (*f*). *g*, A *hh*, HS-*wg* embryo, double labelled for anti En and anti β -galactosidase in which En is induced ectopically. In this case the lack of β -galactosidase was used as a marker for the double-mutant embryos. *h*, A *wg*^{cx4};HS-*wg*/+ embryo, labelled for both En and Wg proteins. The absence of Wg protein expression indicates that this embryo is homozygous for *wg*^{cx4}, but it does show En expansion as observed in HS-*wg*. *i*, Subcellular localization of En protein in HS-*wg* (*i*) and *en*^{cx1};HS-*wg*/+ embryos (*j*). In HS-*wg* En expression is localized to the nucleus and in a broader domain than in wild type (*a*). In *en*^{cx1};HS-*wg*/+ the En domain is broadened but the protein is localized to the cytoplasm as seen in *en*^{cx1} embryos¹¹.

METHODS. The HS-*wg*/TM3, *Sb* stock has been described previously¹⁷. Several segment polarity mutations used here have previously been characterized as strong alleles: *en*^{cx1} (ref. 11), *wg*^{cx4} (ref. 3), *hh*^{G51} (ref. 19). From the mutant alleles *arm*^{25B}FRT¹⁰¹, *arm*^{KM19}FRT¹⁰¹ (ref. 23), *dsh*⁴⁷⁷FRT¹⁰¹, *dsh*⁷⁵FRT¹⁰¹ (N.P., unpublished) and *dsh*^{V26}FRT¹⁰¹ (ref. 15), mosaic germ lines were made by the FLP-DFS technique³¹, whereas from *porc*^{PB16} germ-line clone embryos were made by X-ray irradiation using the DFS technique³². To study the HS-*wg* phenotype in a segment polarity mutant background, we generated embryos that contain HS-*wg* and lack one of the following segment polarity genes: *hh*, *en*, *wg*, *dsh*, *arm* and *porc*. Because HS-*wg* flies are very weak, it was in most cases not possible to establish stocks that contain the HS-*wg* P-element over a TM3 balancer chromosome and a balanced segment polarity mutation. A recombinant stock was obtained for *hh* and HS-*wg* on the third chromosome and balanced over a TM3 chromosome, marked with a

P-element carrying the β -galactosidase gene under the control of the *hunchback* promoter (G. Struhl, unpublished results). Using this marker we could identify embryos that are homozygous for the HS-*wg* P-element and *hh*^{G51}. In the case of the second chromosome mutations *en*^{cx1} and *wg*^{cx4}, crosses were made between HS-*wg*/TM3 males and *en*^{cx1}/Gla or *wg*^{cx4}*b pr*/CyO females. F₁ males and females, carrying the particular segment-polarity mutation and one copy of the HS-*wg* P-element, were selected and F₂ progeny collected and heat-shocked. In a double-antibody staining of WG and EN proteins in *wg*^{cx4};HS-*wg* embryos and a single En staining in *en*^{cx1};HS-*wg* embryos were identified that showed a novel staining pattern not present in the mutant or the HS-*wg* embryos alone. In this way we were able to identify the double mutant embryos unambiguously. Females carrying homozygous germ-line clones of *dsh*, *arm* and *porc* were crossed with the *FM7*, *ftz LacZ*/Y;HS-*wg*/+ males. The *FM7*, *ftz LacZ* balancer chromosome marked with the β -galactosidase gene under the control of the *fushi tarazu* promoter³³ allows marking of the embryos from about stage 6 to 13. From this cross, stage 6-13 embryos that do not stain with anti- β -galactosidase antibody lack the particular segment polarity gene on the first chromosome. Half of these embryos carry the HS-*wg* P-element. By double labelling with anti- β -galactosidase and anti-*wg* or anti-*en* antibodies the pattern of Wg and En proteins in the double mutant embryos could be determined. The maximal broadening of En in stage 10 to 12 (ref. 18) HS-*wg* embryos was only seen after multiple heat-shocks¹⁷. As a consequence of the heat-shock procedure, generalized HS-*wg* protein is present in these embryos starting at around 2 h until 6.5 h AEL (stages 4 to 10). Fixation of embryos and double labelling for En and Wg or for En and β -galactosidase (Promega, Cappel) were as described¹⁷. For antibody stainings and cuticle preparations (Fig. 2) the appropriate heat-shocked segment polarity mutant embryos were used as a control for the general effects of heat shock.

FIG. 2 Effects of lack of *dsh*, *arm*, *porc*, *hh* or endogenous *wg* on cuticle phenotype of HS-*wg* embryos. Anterior is up. All embryos used for cuticle preparations were heat-shocked as described¹⁷. Cuticles were mounted in Hoyer's and photographed under dark-field optics at the same magnification. *a*, Ventral view of wild-type larvae. The posterior spiracle and associated Filzkörper material is indicated by the arrow. *b*, Ventral cuticle pattern of an HS-*wg* embryo. No ventral denticles are present, except for the beard¹⁷. *c*, A *wg^{cx4}* embryo, showing a lawn of disoriented denticles. Head structures are abnormal and filzkörper are missing. *d*, A *dsh⁴⁷⁷*;HS-*wg*/+ embryo. These embryos are indistinguishable from *dsh* and *wg* with the exception that in some larvae some rudimentary filzkörper material is present. *e*, An *arm^{25b}*;HS-*wg*/+ embryo. These embryos are indistinguishable from *arm* and *wg* embryos. *f*, A *porc^{PB16}*;HS-*wg*/+ embryo. These embryos show a slight restoration of pattern and naked cuticle. In addition, the posterior spiracles with their filzkörper materials are present (Not shown). *g*, Ventral view of a *hh^{G51}*, HS-*wg* embryo. Some patches of denticles are left, but most of cuticle is naked except for the beard (indicated by the arrow). *h*, Cuticle pattern of a *wg^{cx4}*;HS-*wg* embryo. A 20-min heat shock between 4 to 9 h AEL substantially rescues segmental pattern, head structures and filzkörper materials (arrow).

METHODS. The most extreme cuticle phenotype in HS-*wg* was generated with a single heat shock between 4 to 9 h of development (stages 8 to 11 (refs 17, 18)). Therefore, to analyse the cuticle phenotypes of the double-mutant larvae, stage 8 to 11 embryos were individually selected and heat-shocked once for 20 min. Because we did not have any markers to distinguish between the single- and double-mutant embryos, we counted the numbers in different progeny classes. We were thus able to determine the phenotype of the double mutant larvae. Larvae were mounted¹⁸ in Hoyer's mountant.

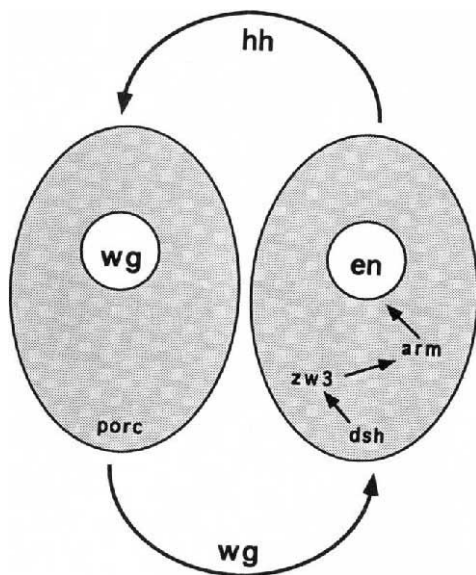
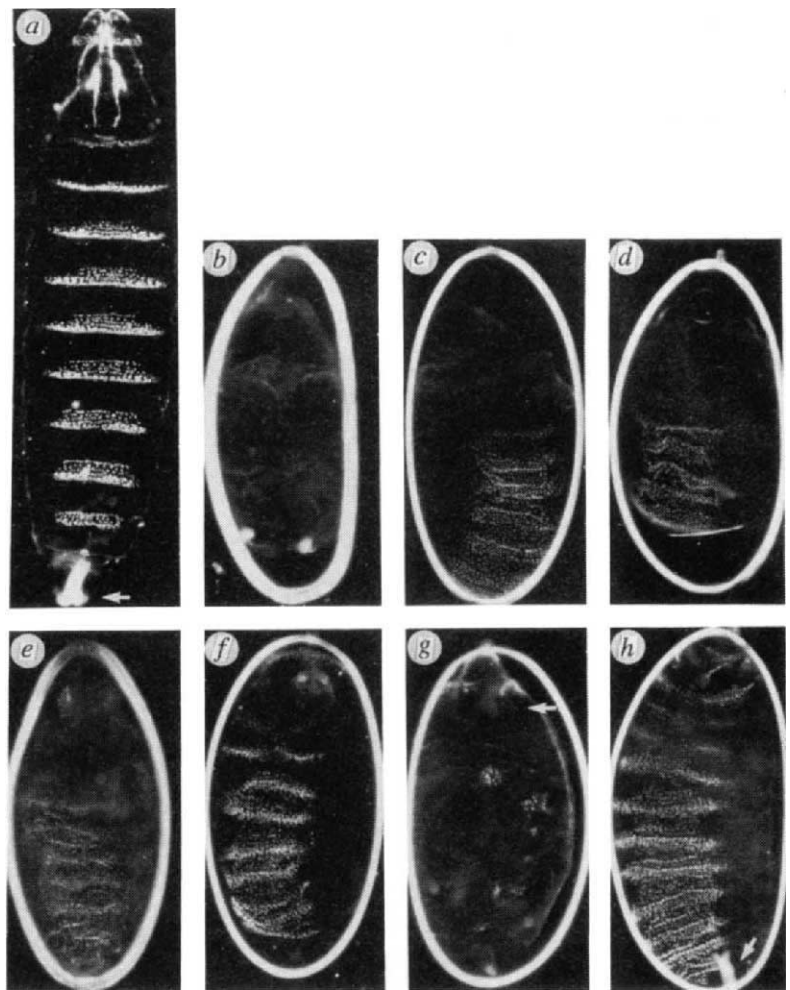


FIG. 3 A model for intercellular signalling between Wg- and En-expressing cells in the embryonic epidermis. Wg protein is secreted^{7,8} and required in the neighbouring cells to maintain the nuclear protein En⁹⁻¹². Secretion of Wg appears to be mediated by the product of the *porc* gene²⁴. *dsh* and *arm* are downstream of *wg* in their effect on *en* expression and act cell-autonomously, indicating that they play a role in reception rather than secretion of Wg. By genetic analysis it has been shown that *zw3* is downstream of *dsh* and upstream of *arm* (ref. 30). *hh* encodes a putatively secreted protein and is transcribed in the En cells²⁵. Its function is upstream of *wg* in our analysis. It has been postulated that the Hh protein antagonizes repression of *wg* expression²⁶ and might be the signal emanating from the En cells to maintain Wg. The arrows indicate genetic relationships; little is known about the biochemistry of these interactions.

downstream of *zw3* (Fig. 3). Because all the cloned genes in the *wg* pathway have homologues in mammals, the mechanism of action of *Wnt* genes during development of highly diverse species might be conserved. □

Received 2 August; accepted 19 October 1993.

- Nusse, R. & Varmus, H. E. *Cell* **69**, 1073-1087 (1992).
- McMahon, A. P. *Trends Genet.* **8**, 236-242 (1992).
- Baker, N. E. *EMBO J.* **6**, 1765-1773 (1987).
- Rijsewijk, F. et al. *Cell* **50**, 649-657 (1987).
- Cabrera, C. V., Alonso, M. C., Johnston, P., Phillips, R. G. & Lawrence, P. A. *Cell* **50**, 659-663 (1987).
- Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
- Van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. A. *Cell* **59**, 739-749 (1989).
- González, F., Swales, L., Bejsovec, A., Skaer, H. & Martínez-Arias, A. *Mech. Dev.* **35**, 43-54 (1991).
- DiNardo, S., Sher, E., Heemskerk-Jorgens, J., Kassisi, J. & O'Farrell, P. *Nature* **332**, 604-609 (1988).
- Martínez-Arias, A., Baker, N. E. & Ingham, P. W. *Development* **103**, 157-70 (1988).
- Heemskerk, J., DiNardo, S., Kostriken, R. & O'Farrell, P. H. *Nature* **352**, 404-410 (1991).
- Bejsovec, A. & Martínez-Arias, A. *Development* **113**, 471-185 (1991).
- Dougan, S. & Dinardo, S. *Nature* **360**, 347-350 (1992).
- Klingensmith, J., Noll, E. & Perrimon, N. *Dev. Biol.* **134**, 130-145 (1989).
- Perrimon, N. & Mahowald, A. P. *Dev. Biol.* **119**, 587-600 (1987).
- Wieschaus, E. & Riggleman, R. *Cell* **49**, 177-84 (1987).
- Noordermeer, J., Johnston, P., Rijsewijk, F., Nusse, R. & Lawrence, P. *Development* **116**, 711-719 (1992).
- Wieschaus, E. & Nüsslein-Volhard, C. in *Drosophila: A Practical Approach* (ed. Roberts, D. B.) (IRL, Oxford, 1986).
- Mohler, J. *Genetics* **120**, 1061-1072 (1988).
- Sampedro, J., Johnston, P. & Lawrence, P. A. *Development* **117**, 677-687 (1993).
- Klingensmith, J., Nusse, R. & Perrimon, N. *Genes Dev.* (in the press).
- Riggleman, B., Schedl, P. & Wieschaus, E. *Cell* **63**, 549-560 (1990).
- Peifer, M. & Wieschaus, E. *Cell* **63**, 1167-1178 (1990).
- Van den Heuvel, M., Harryman-Samos, C., Klingensmith, J., Perrimon, N. & Nusse, R. *EMBO J.* (in the press).
- Lee, J. J., von Kessler, D. P., Parks, S. & Beachy, P. A. *Cell* **71**, 33-50 (1992).
- Ingham, P. W., Taylor, A. M. & Nakano, Y. *Nature* **353**, 184-187 (1991).
- Bourouis, M. et al. *EMBO J.* **9**, 2877-2884 (1990).

28. Siegfried, E., Perkins, L. A., Capaci, T. M. & Perrimon, N. *Nature* **345**, 825–829 (1990).
29. Siegfried, E., Chou, T. B. & Perrimon, N. *Cell* **71**, 1167–1179 (1992).
30. Siegfried, E., Wilder, E. L. & Perrimon, N. *Nature* **367**, 76–80 (1994).
31. Chou, T. B. & Perrimon, N. *Genetics* **131**, 643–653 (1992).
32. Eberl, D. F., Perkins, L. A., Engelstein, M., Hiliiker, A. J. & Perrimon, N. *Genetics* **130**, 569–583 (1992).
33. Kania, M. A., Bonner, A. S., Duffy, J. B. & Gergen, J. P. *Genes Dev.* **4**, 1701–1713 (1990).

ACKNOWLEDGEMENTS. We thank G. Barsh, M. Scott, E. Siegfried and the members of the Nusse laboratory for comments on the manuscript and L. Fradkin for discussions, R. Holmgren, T. Kornberg, J. Mohler, C. Nüsslein-Volhard and G. Struhl for fly stocks, C. Girdman and C. Goodman for anti-En, and K. Shaw for secretarial assistance. *arm^{25b}FRT¹⁰¹* and *arm^{TM19}FRT¹⁰¹* recombinant chromosome stocks were provided by E. Wieschaus. This work was supported by the Howard Hughes Medical Institute. N.P. and R.N. are investigators of the Howard Hughes Medical Institute.

Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different cis-regulatory regions

Xuelin Li & Markus Noll

Institute for Molecular Biology II, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

It is generally accepted that the specific function of a gene depends on its coding sequence. The three paired-box and homeobox genes *paired* (*prd*), *gooseberry* (*gsb*) and *gooseberry neuro* (*gsbn*) have distinct developmental functions in *Drosophila* embryogenesis^{1–5}. During the syncytial blastoderm stage, the pair-rule gene *prd*^{4,6} activates segment-polarity genes, such as *gsb*⁷, *wingless* (*wg*), and *engrailed* (*en*), in segmentally repeated stripes⁸. After germ-band extension, *gsb* maintains the expression of *wg*, which in turn specifies the denticle pattern by repressing a default state of ubiquitous denticle formation in the ventral epidermis⁹. In addition, *gsb* activates *gsbn*⁵, which is expressed mainly in the central nervous system^{2,3}, suggesting that *gsbn* is involved in neural development. Here we show that, despite the functional difference and the considerably diverged coding sequence of these genes, their proteins have conserved the same function. The finding that the essential difference between genes may reside in their cis-regulatory regions exemplifies an important evolutionary mechanism of how function diversifies after gene duplication.

The most conspicuous feature of the segmental organization of a *Drosophila* larva is its ventral denticle pattern (Fig. 1a). Recently, we have shown that *gsb* regulates this pattern through a *wg*–*gsb* autoregulatory loop that maintains the expression of *wg*, which represses denticle formation⁹. Thus, when *Hsgsb* embryos carrying a transgenic *gsb* gene under the control of the heat-inducible *hsp70* promoter were heat-shocked between 3 h 10 min and 6 h 20 min of development at 25 °C (early time interval), ubiquitous expression of *gsb* generated a naked larval cuticle (Fig. 1e). An earlier heat shock between 2 h 10 min and 3 h 10 min of development at 25 °C (early time interval), however, induced a pair-rule phenotype (Fig. 1b). This result is unexpected because it differs dramatically from the normal *gsb* gain-of-function phenotype (Fig. 1e). In wild-type embryos, *gsb* begins to be expressed only by the end of this early time interval, which coincides with the time of pair-rule gene rather than segment-polarity gene function. In fact, ubiquitous expression of pair-rule genes is known to result in pair-rule phenotypes that are nearly reciprocal to their loss-of-function phenotypes^{10–14}. Therefore, we suspected that activation of *Hsgsb* during the early time interval mimics the function of a ubiquitously expressed pair-rule protein. The most likely candidate was the Prd protein, as its N-terminal half consists of a paired-domain and a *prd*-

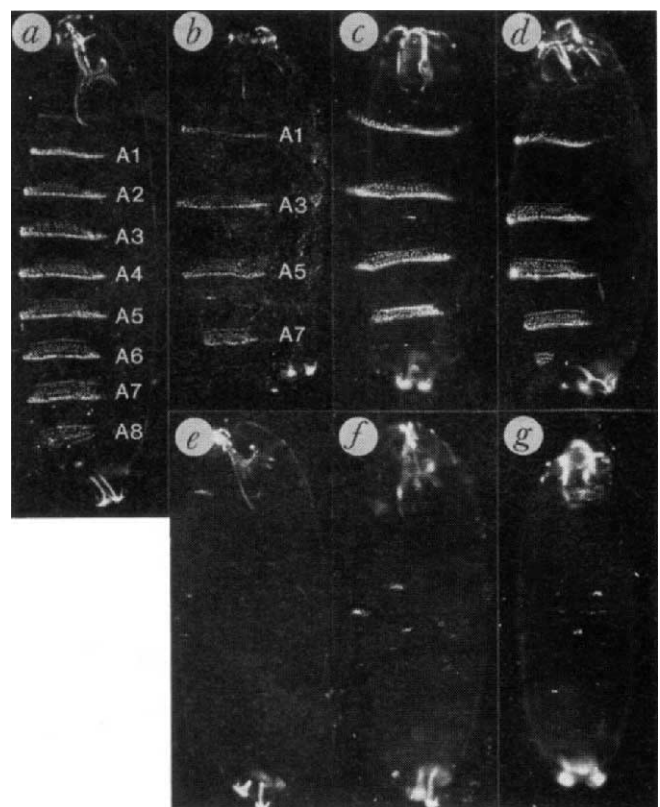


FIG. 1 Identical cuticular phenotypes induced by the ubiquitous expression of *gsb*, *prd* or *gsbn*. Cuticular preparations of wild-type (a), *Hsgsb* (b, e), *Hsprd* (c, f) and *Hsgsbn* (d, g) embryos heat-shocked during the early (a–d) or late (e–g) time interval are shown as ventral views under dark-field illumination. Ubiquitous activation of *Hsgsb*, *Hsprd* or *Hsgsbn* during the early time interval generates a pair-rule cuticular phenotype. In all cases, even-numbered abdominal denticle belts (A2, A4, A6, A8) and their anteriorly adjacent naked regions are lost, with the occasional exception of a few remaining denticles. This phenotype is nearly reciprocal to that of *prd*[−] embryos in which the odd-numbered denticle belts and their anteriorly neighbouring naked regions are deleted. Ubiquitous activation of *Hsgsb*, *Hsprd* and *Hsgsbn* by a heat shock during the late time interval produces a naked cuticular phenotype.

METHODS. Transgenic *Hsgsb*, *Hsprd* or *Hsgsbn* embryos, collected between 2 h 10 min and 3 h 10 min AEL (after egg laying) (early time interval) or between 3 h 10 min and 6 h 20 min AEL (late time interval), were heat-shocked for 15 min at 37 °C. After 24 h of development at 25 °C, cuticles were prepared essentially as described²⁸. Transgenic *Hsprd*, *Hsgsb* and *Hsgsbn* fly stocks were produced, as previously described²⁹, by cloning a *prd*-cDNA, c7340.4 (ref. 19), a *gsb*-cDNA, BSH9c2, or a *gsbn*-cDNA, BSH4c4 (ref. 3), into the P-element vector pKB255 (K. Basler and E. Hafen, unpublished) and subsequent germ-line transformation of *w¹¹¹⁸* embryos according to standard procedures³⁰.

type homeodomain and thus is highly homologous to the N-terminal half of Gsb². Indeed, early ubiquitous expression of *prd* in *Hsprd* embryos produced a phenotype¹² (Fig. 1c) indistinguishable from the *Hsgsb* pair-rule phenotype (Fig. 1b).

As *gsb* maintains the expression of *wg*⁹, we expect that the pair-rule phenotype of *Hsgsb* and *Hsprd* results from ectopic expression of the endogenous *gsb* and *wg* genes. Indeed, ubiquitous activation of either *gsb* or *prd* during the early time interval generates ectopic Gsb (Fig. 2b, c) and Wg stripes (Fig. 2f, g) anterior to the even-numbered wild-type Gsb (Fig. 2a) and corresponding Wg stripes (Fig. 2e). The observed pair-rule phenotype (Fig. 1b, c) is thus consistent with the ectopic *wg* expression and the resulting repression of denticle formation (Fig. 1b, c).

Activation of *Hsprd* has been shown to expand the odd-num-

bered En stripes posteriorly¹² (Fig. 2i, l). We now show that the same effect occurs in heat-shocked *Hsgsb* embryos (Fig. 2k). Note that the initial ectopic activation of *gsb*, *wg* and *en* by *Hsprd* or *Hsgsb* depends on the pair-rule function of Prd, which activates the segmental stripes of Gsb, Wg and En^{8,12}. In contrast, the subsequent maintenance of this ectopic expression of *gsb*, *wg* and *en* (Fig. 2a–m) is determined solely by their mutual activations and hence becomes independent of the product of the *Hsprd* or *Hsgsb* transgene.

Therefore, after heat shock within the early time interval, both *Hsgsb* and *Hsprd* embryos exhibit the same altered expression patterns of the endogenous *gsb*, *wg* and *en* genes and develop indistinguishable pair-rule phenotypes. We conclude that at this

time *Hsgsb* functions as *Hsprd*.

Just as Gsb can substitute for Prd during the early time interval, so might Prd replace Gsb function during the late time interval. Indeed, when we examined whether heat-induced *prd* expression during the late time interval repressed denticle formation in *Hsprd* embryos, we found that denticles are strongly repressed (Fig. 1f) as in the case of *Hsgsb* embryos treated similarly (Fig. 1e). Yet none of the embryos exhibit the pair-rule phenotype observed after heat shock during the early time interval.

Consistent with the altered phenotype observed after late activation of *Hsgsb* or *Hsprd*, expression of the endogenous *gsb*, *wg* and *en* genes shows a different response from that after early

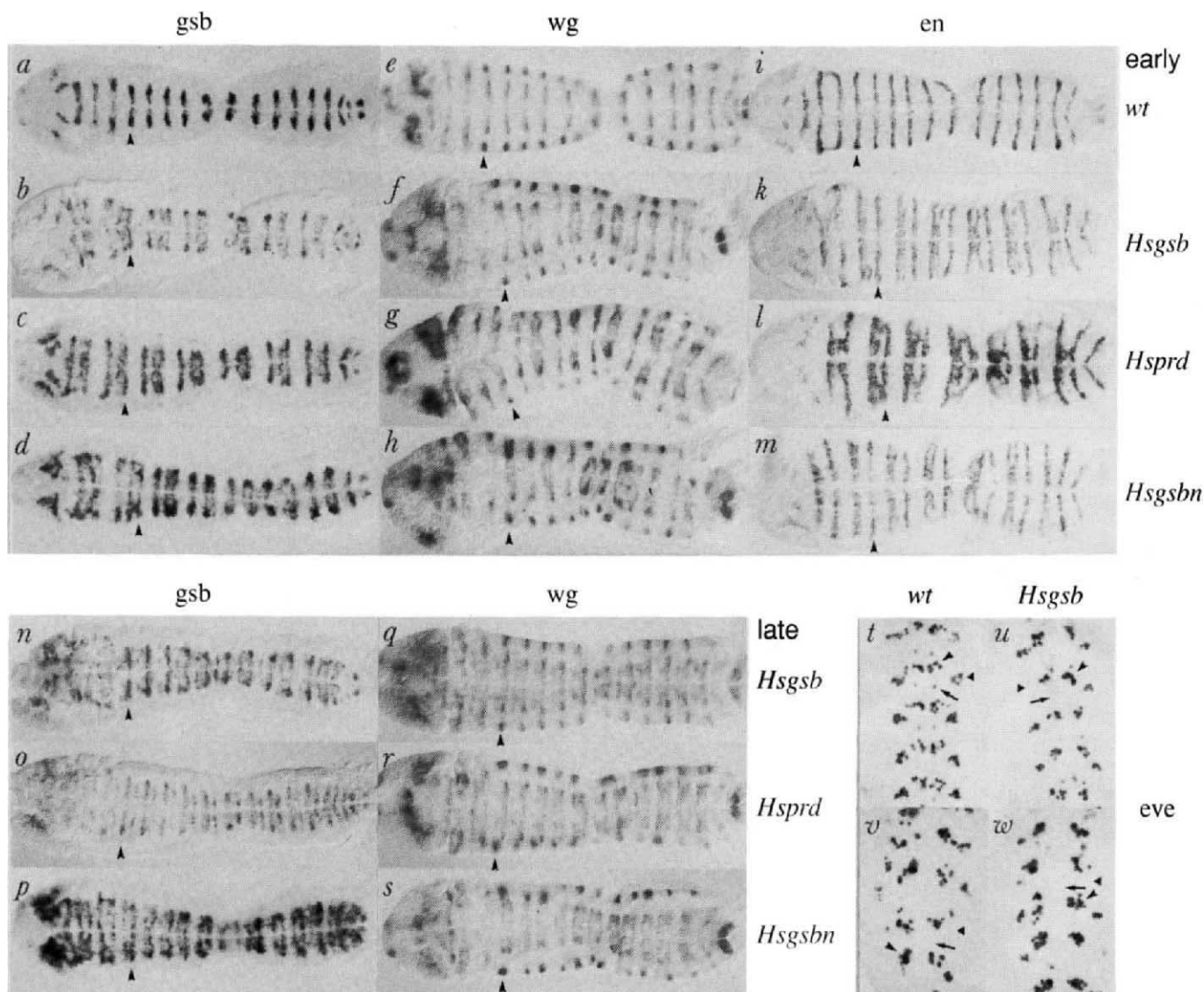


FIG. 2 Ubiquitous expression of *gsb*, *prd* or *gsbn* alters the expression of *gsb*, *wg*, *en* and *eve* in the same manner. a–s, Ectopic expression of *gsb*, *wg* and *en* induced by the ubiquitous expression of *gsb*, *prd* or *gsbn* during the early and late time interval. Expression patterns of *gsb* (a–d, n–p), *wg* (e–h, q–s) and *en* (i–m) are shown in wild-type (a, e, i), *Hsgsb* (b, f, k, n, q), *Hsprd* (c, g, l, o, r) and *Hsgsbn* (d, h, m, p, s) embryos 3–4 h after early (a–m) or late (n–s) heat-shock treatment. Embryos, oriented with their anterior to the left, are cut and unfolded along the amnioserosa to show the entire set of stripes. Arrowheads point to stripe 4 of Gsb, Wg or En. For convenience, the numbering of the Wg stripes follows that of the corresponding Gsb and En stripes^{3,31}. Thus, stripe 4 of Wg and En are adjacent to each other³² whereas stripe 4 of *gsb* overlaps with stripe 4 of both Wg and En (ref. 5). t–w, Change of *eve* expression in the CNS induced by the ubiquitous expression of

gsb, *prd* or *gsbn* during the late time interval. The expression in the CNS of *eve* is shown in wild-type (t), *Hsgsb* (u), *Hsprd* (v) and *Hsgsbn* (w) embryos 10 h after late heat shock. A trunk region of the CNS is shown with its anterior oriented up. Note the frequent loss of RP2 (arrows) and EL (triangles) neurons and the amplified CQ neurons (arrowheads) in *Hsgsb*, *Hsprd* or *Hsgsbn* compared with wild-type embryos.

METHODS. Transgenic *Hsgsb*, *Hsprd* or *Hsgsbn* embryos, collected between 2 h 30 min and 3 h 10 min AEL (a–m), between 3 h 40 min and 4 h 20 min AEL (n–s), or between 4 h 30 min and 6 h 30 min AEL (t–w) at 25 °C, were heat-shocked for 15 min at 37 °C and allowed to recover for 4 h (a–m), 3 h (n–s) or 10 h (t–w) at 25 °C before fixation and staining with anti-Gsb, anti-Wg, anti-En or anti-Eve antibodies as described¹⁸.

heat induction. Although heat shock within the late time interval induces an ectopic *gsb* and *wg* stripe anterior to each wild-type stripe in both *Hsgsb* (Fig. 2n, q) and *Hsprd* (Fig. 2o, r) embryos, ectopic *en* expression is no longer observed (not shown). As *gsb* does not affect *en* expression⁴, these observations imply that, during the late time interval, *prd* expressed ectopically carries out only the segment-polarity function of the ectopic Gsb protein.

The differences in cuticular phenotypes and expression of *gsb*, *wg* and *en* between the early and late activation of *Hsprd* or *Hsgsb* further suggest the existence of a dramatic transition from a 'pair-rule' stage to a 'segment-polarity' stage at about 3 h 10 min during *Drosophila* embryogenesis.

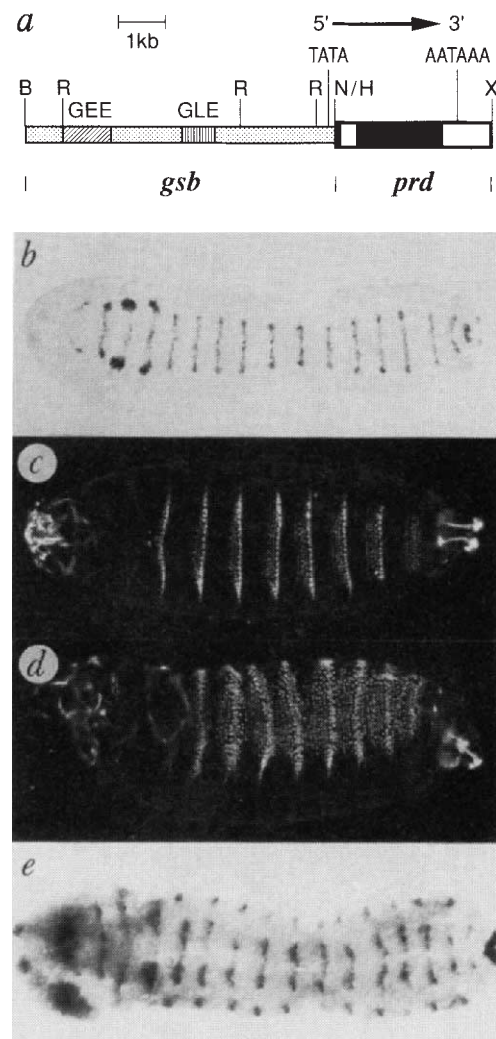
The *gsbn* gene encodes a protein whose N terminus, like that of the products of *gsb* and *prd*, consists of a paired-domain and a *prd*-type homeodomain but whose C terminus differs from those of the other two²³. Therefore, we investigated whether *Hsgsbn* could also substitute for *Hsgsb* or *Hsprd*, and found that indeed, heat-activated *Hsgsbn* produces the same pair-rule and naked cuticular phenotypes as *Hsgsb* and *Hsprd* (Fig. 1d, g). Moreover, activation of *Hsgsbn* during the early and late time interval again changes the expression patterns of *gsb*, *wg* and *en* in the same way as the similarly induced *Hsgsb* and *Hsprd* (Fig. 2a–s). Therefore Prd, Gsb and Gsbn can all replace each other with respect to these criteria.

To compare the effects of ubiquitous expression of *prd*, *gsb* or *gsbn* on central nervous system (CNS) development, *Hsprd*, *Hsgsb* or *Hsgsbn* embryos were heat-shocked during the late time interval and immunologically stained for the Even-skipped (Eve) protein. In addition to its role in segmentation, the pair-rule gene *eve* is expressed in and specifies the fate of certain neurons in the CNS^{15–17}. As expected, heat induction of *Hsprd*, *Hsgsb* or *Hsgsbn* disturbs the stereotype wild-type *eve* expression pattern in the CNS in a similar way (Fig. 2t–w). Particularly striking are the frequent loss of the RP2 and EL neurons and the increased number of the CQ neurons, suggesting that expression of *Hsprd*, *Hsgsb* and *Hsgsbn* interferes similarly with neural development.

Although our results show that the Prd protein can replace Gsb in its ectopic function, it is important to test whether it can also substitute for normal Gsb function. Therefore, we attempted to rescue the cuticular phenotype of *gsb*[−] embryos by expressing a *prd* transgene in cells that normally express *gsb*. The *gsb* control region, including the *cis*-regulatory elements for the segmental stripes of *gsb*-ectodermal expression¹⁸, was fused to the *prd* gene comprising the entire coding region¹⁹ (Fig. 3a) and used to generate transgenic flies. As expected, embryos express this *gsb*-*prd* transgene in a *gsb*-like pattern (Fig. 3b). Moreover, the *gsb* cuticular phenotype of homozygous

FIG. 3 Rescue of the *gsb*[−] cuticular phenotype by Prd protein. **a**, Structure of the *gsb*-*prd* transgene. The 6.6-kilobase (kb) *Bam*HI–*Nru*I fragment of *gsb*, containing the upstream regulatory elements GEE and GLE, the TATA box and 104 base pairs (bp) of the *gsb* leader sequence¹⁸, was fused to the 3.2-kb *Hind*III–*Xba*I genomic fragment of *prd*, comprising 32 bp of the leader, the coding region, including the intron, and the entire trailer sequence¹⁹. GEE and GLE are the control elements responsible for the establishment and maintenance of *gsb* expression in the ectoderm¹⁸. The *prd* leader and coding region are shown as black boxes, and open boxes indicate the *prd* intron, the 3' trailer and part of the downstream region. Abbreviations of restriction sites: B, *Bam*HI; H, *Hind*III; N, *Nru*I; R, *Eco*RI; X, *Xba*I. **b**, Expression pattern of the *gsb*-*prd* transgene in embryos stained with anti-Prd antibodies. The embryo, oriented with its anterior to the left, is at stage 11 (5.5 h AEL) and has been unfolded to show the entire set of stripes. At this time, the endogenous *prd* protein is no longer detectable⁴. The stained Prd protein originates from the *gsb*-*prd* transgene and shows a similar pattern to that of the endogenous Gsb protein (see Fig. 2a). **c**, **d**, Rescue of the *gsb*[−] cuticular phenotype by the *gsb*-*prd* transgene. Ventral views of the cuticle preparations of homozygous *Df*(2R)*gsb*^{1X62} embryos with (c) or without (d) a *gsb*-*prd* transgene are shown under dark-field illumination. As the deficiency *Df*(2R)*gsb*^{1X62} deletes in addition to *gsb* and *gsbn* at least another lethal gene, *zipper*²⁰, lethality is not rescuable. Note that the head defects resulting from the *zipper* mutation are visible in both embryos (oriented with their anterior to the left). **e**, Activation of *wg* by the *gsb*-*prd* transgene in homozygous *Df*(2R)*gsb*^{1X62} embryos. Embryos have been stained with anti-Wg and anti-Gsbn antibodies to identify those embryos that fail to stain for Gsbn and thus are homozygous for *Df*(2R)*gsb*^{1X62}. The embryo shown, unfolded and oriented with its anterior to the left, is at the late stage 11 (7 h AEL). Note that at this stage *wg* expression has completely decayed in homozygous *Df*(2R)*gsb*^{1X62} embryos that carry no *gsb*-*prd* transgene^{9,33}.

METHODS. To prepare the *gsb*-*prd* rescue construct, the 1.5-kb *Nru*I–*Sma*I fragment (+104 bp to +1.6 kb) of the plasmid 9Z2¹, which contains the 8.1-kb *Bam*HI fragment of *gsb* (−6.5 kb to +1.6 kb) in pKSpL3 (ref. 18), was replaced by the blunt-ended 3.6-kb *Hind*III–*Xho*I fragment of *prd* from the genomic DNA clone D7.11 (ref. 6), generating the *gsb*-*prd* gene in pKSpL3. From this construct, the *gsb*-*prd* gene was removed as *Xba*I fragment (one *Xba*I site in polylinker, the other ~400 bp upstream of the *Xho*I site of *prd*) and cloned into the *Xba*I site of the pW6 vector containing the P-element and the mini-white gene³⁴. The resulting *gsb*-*prd* rescue construct was injected into *w*¹¹¹⁸ embryos as described³⁰, and 11 transgenic lines were obtained. The transgenic *gsb*-*prd* embryos were stained with anti-Prd antibodies⁴ to verify that the *gsb*-*prd* gene is expressed in a *gsb*-like pattern. To test whether *gsb*-*prd* can rescue the *Df*(2R)*gsb*^{1X62} cuticular phenotype, six transgenic lines, carrying the *gsb*-*prd* transgene on either the first or third chromosome, were crossed with *w*¹¹¹⁸; *Df*(2R)*gsb*^{1X62}/CyO flies to gen-



erate *Df*(2R)*gsb*^{1X62}/+; *gsb*-*prd*/(+ or *w*¹¹¹⁸) flies. Embryos from these flies were allowed to develop until 24 h AEL, when cuticles were prepared as described²⁸.

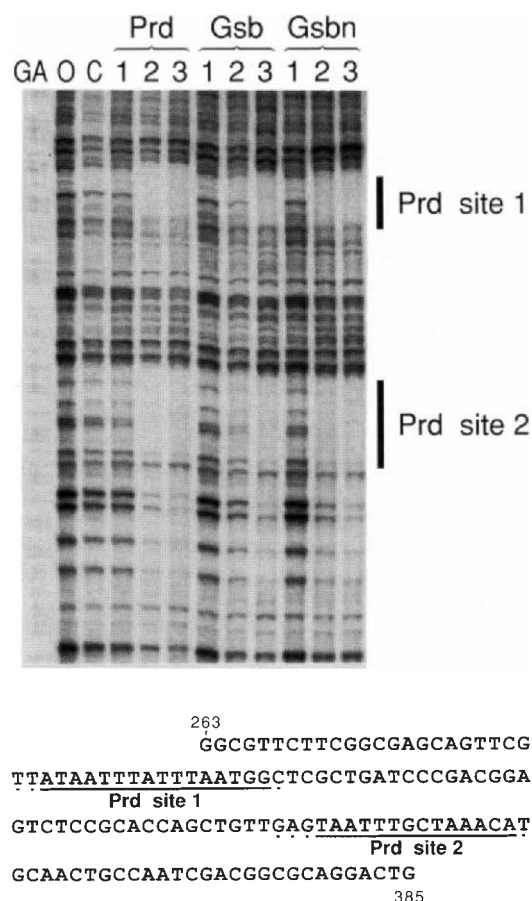


FIG. 4 Identical DNA-binding specificities of *prd*, *gsb* and *gsbn* proteins as analysed by DNase I footprinting. A 510-bp *EcoRI*-*Bgl*III fragment from GEE¹⁸, 3' end-labelled at the *Bgl*III site with [α -³²P]-dATP and Klenow enzyme, was incubated with 5, 20 or 50 µg (lanes 1–3) of Prd, Gsb or Gsbn protein extracts. The same two sites were protected by Prd, Gsb and Gsbn protein against DNase I digestion in the region whose sequence is shown below the autoradiograph and extends between 263 bp and 385 bp downstream of the *EcoRI* site located 5.7 kb upstream from the *gsb* transcriptional start site¹⁸. Lane O: control without added protein. Lane C: control with 50 µg of protein extract from cells containing the empty expression vector pAR3040. Lane GA: corresponding G + A sequence³⁵.

METHODS. pARGsb and pARGsbn plasmids were constructed by creating a *Nde*I site at the translation initiation site of the *gsb* and *gsbn* coding sequences using polymerase chain reaction mutagenesis as described²², and subsequent cloning of each sequence into the unique *Nde*I site of the T7 expression vector pAR3040. *Escherichia coli* BL21 cells containing pARprd²², pARGsb, pARGsbn or pAR3040 were induced at a density of $A_{600} \approx 0.5$ with 0.5 mM isopropylthiogalactoside for 2 h. Induced cells were centrifuged and resuspended in 1/100 volume of 50 mM NaCl, 10 mM HEPES, pH 7.6, sonicated and recentrifuged. After another cycle of resuspension, sonication and centrifugation, the insoluble fraction was dissolved in 1/400 volume of 8 M urea, buffer Z (100 mM KCl, 25 mM HEPES, pH 7.6, 12.5 mM MgCl₂, 10 µM ZnSO₄, 1 mM dithiothreitol, 0.1% NP-40, 20% glycerol) and sonicated. The soluble fraction contains the highly concentrated induced proteins that constitute most of its total protein. As the induced proteins are very insoluble without urea, small aliquots of these extracts are used directly in DNase I footprint assays as described³⁶. After DNase I digestion, the DNA is run on a 8% polyacrylamide/7.5 M urea gel and analysed by autoradiography.

Df(2R)gsb^{1X62} embryos (Fig. 3d), in which both *gsb* and *gsbn* are deleted^{2,20}, is completely rescued by the *gsb-prd* transgene of all six independent transgenic lines tested (Fig. 3c). Because in wild-type embryos, *gsb* maintains the expression of *wg*, which represses denticle formation^{9,21}, we expect, and find, that the rescue of the *gsb*⁻ cuticular phenotype results from activation of *wg* by *gsb-prd* in *gsb*⁻ embryos (Fig. 3e). These results corroborate our previous conclusion that Prd can substitute for the function of Gsb.

The three proteins Prd, Gsb and Gsbn are transcription factors because they contain a homeo- and a paired-domain. Because they can replace each other's gene regulatory functions, we expect them to recognize the same DNA sequences. To examine their DNA-binding specificities, each protein was analysed by footprinting on a previously characterized *gsb* cis-regulatory element, GEE. This element activates *gsb* by pair-rule proteins including Prd¹⁸, and so should be a target for Prd protein binding *in vivo*. As shown in Fig. 4, all three proteins bind precisely to the same sites, demonstrating that they share the same DNA-binding specificity. It is unclear whether the paired-domain, homeodomain or both participate in the binding to these extremely (A+T)-rich sequences. Their similarity to previously defined homeodomain-binding sites²², however, suggests an involvement of the homeodomain in this interaction.

It was startling to discover that Prd, Gsb and Gsbn, very different in their C-terminal halves, are functionally equivalent, as this implies that the essential determinant of the function of the three genes is encoded in their specific cis-regulatory regions controlling their temporal and spatial expression. As the N-terminal paired- and homeodomains have been highly conserved in these genes, they must have evolved from the same ancestral gene by repeated duplication. Our results thus provide the first experimental example that, during evolution, duplicated genes

may acquire new functions by changes in their regulatory regions generating an altered expression, rather than by mutations in their coding sequences. Moreover, the highly diverged C-terminal halves of Prd, Gsb and Gsbn imply that considerable changes in the coding region may be tolerated without significantly altering the function of the protein. The idea that mutations in the cis-regulatory rather than coding regions drive functional diversification has already been suggested²³.

We propose that during duplication, the duplicated portion of a gene, including its coding region, is juxtaposed to new cis-regulatory elements which change its expression and thus give it a new function—ultimately the activation of a different set of genes. Subsequently, mutations accumulating in the coding region may further contribute to an altered function of the gene. This hypothesis provides a simple explanation for what seem to be redundant functions revealed in gene knock-outs in vertebrates^{24–27}. In these cases, elimination of a gene's function is partly compensated for by the overlapping expression of a homologous protein with an equivalent function. If during evolution a new gene arises by the combination of new cis-regulatory sequences with the duplicated coding portion and some of the cis-regulatory elements of an old gene, the expression of the new gene will overlap in time and space with that of the old gene and generate redundancy. □

Received 4 August; accepted 27 October 1993.

1. Noll, M. *Curr. Opin. Genet. Dev.* **3**, 595–605 (1993).
2. Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. *Cell* **47**, 1033–1040 (1986).
3. Baumgartner, S., Bopp, D., Burri, M. & Noll, M. *Genes Dev.* **1**, 1247–1267 (1987).
4. Gutjahr, T., Frei, E. & Noll, M. *Development* **117**, 609–623 (1993).
5. Gutjahr, T., Patel, N. H., Li, X., Goodman, C. S. & Noll, M. *Development* **118**, 21–31 (1993).
6. Kilchherr, F., Baumgartner, S., Bopp, D., Frei, E. & Noll, M. *Nature* **321**, 493–499 (1986).
7. Bopp, D., Jamet, E., Baumgartner, S., Burri, M. & Noll, M. *EMBO J.* **8**, 3447–3457 (1989).
8. Ingham, P. W. & Martínez-Arias, A. *Cell* **68**, 221–235 (1992).
9. Li, X. & Noll, M. *EMBO J.* **12**, 4499–4509 (1993).
10. Struhl, G. *Nature* **318**, 677–680 (1985).
11. Ish-Horowitz, D. & Pinchin, S. M. *Cell* **51**, 405–415 (1987).
12. Morrissey, D., Askew, D., Raj, L. & Weir, M. *Genes Dev.* **5**, 1684–1696 (1991).
13. Manoukian, A. S. & Krause, H. M. *Genes Dev.* **6**, 1740–1751 (1992).
14. Manoukian, A. S. & Krause, H. M. *Development* **118**, 785–796 (1993).
15. Frasch, M., Hoey, T., Rushlow, C., Doyle, H. & Levine, M. *EMBO J.* **6**, 749–759 (1987).
16. Doe, C. Q., Smouse, D. & Goodman, C. S. *Nature* **333**, 376–378 (1988).
17. Patel, N. H., Schafer, B., Goodman, C. S. & Holmgren, R. *Genes Dev.* **3**, 890–904 (1989).
18. Li, X., Gutjahr, T. & Noll, M. *EMBO J.* **12**, 1427–1436 (1993).

19. Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. *Cell* **47**, 735–746 (1986).
20. Côté, S. et al. *EMBO J.* **6**, 2793–2801 (1987).
21. Bejsovec, A. & Martínez-Arias, A. *Development* **113**, 471–485 (1991).
22. Hoey, T. & Levine, M. *Nature* **332**, 858–861 (1988).
23. Jacob, F. *Science* **196**, 1161–1166 (1977).
24. Thomas, K. R. & Capocchi, M. R. *Nature* **346**, 847–850 (1990).
25. McMahon, A. P. & Bradley, A. *Cell* **62**, 1073–1085 (1990).
26. Joyner, A. L., Herrup, K., Auerbach, B. A., Davis, C. A. & Rossant, J. *Science* **251**, 1239–1243 (1991).
27. Lohnes, D. et al. *Cell* **73**, 643–658 (1993).
28. Wieschaus, E. & Nüsslein-Volhard, C. in *Drosophila, a Practical Approach* (ed. Roberts, D. B.), 199–227 (IRL Press, Oxford, 1986).
29. Dambly-Chaudière, C. et al. *Cell* **69**, 159–172 (1992).
30. Rubin, G. M. & Spradling, A. C. *Science* **218**, 348–353 (1982).
31. DiNardo, S. & O'Farrell, P. H. *Genes Dev.* **1**, 1212–1225 (1987).
32. van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. A. *Cell* **59**, 739–749 (1989).
33. Hidalgo, A. *Mech. Dev.* **35**, 77–87 (1991).
34. Klemenz, R., Weber, U. & Gehring, W. J. *Nucleic Acids Res.* **15**, 3947–3959 (1987).
35. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
36. Stanoev, D., Hoey, T. & Levine, M. *Nature* **341**, 331–335 (1989).

ACKNOWLEDGEMENTS. We thank K. Basler and E. Hafen for generating the transgenic *Hspsd*, *Hspsb* and *Hspsbn* flies, F. Ochsenbein for photography and artwork, T. Gütjahr for the anti-Gsb antiserum, R. Smith, M. Wilcox, K. Coleman, T. Kornberg, M. van den Heuvel, R. Nusse and A. Martínez-Arias for gifts of anti-En monoclonal antibody and anti-Wg antisera, and H. Noll for discussions and comments on the manuscript. This work has been supported by the Swiss NSF and by the Kanton Zürich.

Transformation by polyoma virus middle T-antigen involves the binding and tyrosine phosphorylation of Shc

Stephen M. Dilworth*, Charlotte E. P. Brewster*, Michael D. Jones†, Luisa Lanfrancone‡, Guilianna Pelicci‡ & Pier G. Pelicci‡

Departments of * Chemical Pathology and † Virology, Royal Postgraduate Medical School, Du Cane Road, London W12 0NN, UK

‡ Istituto di Clinica Medica I, University of Perugia, Via Brunamonti, 06100 Perugia, Italy

POLYOMA virus middle T-antigen converts normal fibroblasts to a fully transformed, tumorigenic phenotype¹. It achieves this, at least in part, by binding and activating one of the non-receptor tyrosine kinases, pp60^{c-src}, pp62^{c-yes} or pp59^{c-fyn} (reviewed in refs 2 and 3). As a result, middle T-antigen itself is phosphorylated on tyrosine residues^{4,5}, one of which (Tyr 315) acts as a binding site for the SH2 domains of phosphatidylinositol-3'-OH kinase 85K subunit^{6–8}. Here we show that another tyrosine phosphorylation site in middle T-antigen (Tyr 250; refs 4, 5) acts as a binding region for the SH2 domain of the transforming protein Shc⁹. This results in Shc also becoming tyrosine-phosphorylated and binding to the SH2 domain of Grb2 (ref. 10). This probably stimulates p21^{ras} activity through the mammalian homologue of the *Drosophila* guanine-nucleotide-exchange factor Sos (reviewed in ref. 11). We suggest that middle T-antigen transforms cells by acting as a functional homologue of an activated tyrosine kinase-associated growth-factor receptor.

Cell transformation by non-receptor tyrosine kinases, such as v-Src or v-Fps, probably involves tyrosine phosphorylation of Shc proteins¹². As middle T-antigen (MT) is known to bind and activate pp60^{c-src}, and this is essential for cell transformation by MT^{2,3}, it is possible that MT uses a similar mechanism. To detect whether an association between MT and Shc exists, immunoprecipitates with one antibody were western-blotted and probed with a second. When no MT was expressed, Shc was only detected in anti-Shc precipitates (Fig. 1A). But with lysates of MT-transformed cells, two polypeptides resembling the 66 and 52K forms (*M_s* 66,000 and 52,000 respectively) of Shc are seen in anti-MT precipitates (Fig. 1B, lane 3). Conversely, a polypeptide migrating with MT is detected in anti-Shc precipitates (lane

2). As the MT and the 52K form of Shc co-migrate on these gels, the experiment was repeated with lysates containing the truncated MT polypeptide encoded by mutant dl8 (ref. 13). Again, two proteins co-migrating with the larger forms of Shc⁹ are seen in anti-MT precipitates (Fig. 1C, lane 3), but a truncated MT species is detected in anti-Shc precipitates (lane 2). Therefore, a specific complex containing MT and the two larger forms of Shc exists in MT-transformed cells.

A series of non-transforming MT mutants (locations are shown schematically in Fig. 2B; their properties are reviewed in refs 14 and 15) were next examined to determine the regions of MT required for Shc binding. A cell line expressing each altered protein was isolated, and the binding properties of the MT polypeptide studied (Fig. 2). None of the MT mutations that disrupt pp60^{c-src} binding (DC1, DC2, NS-2 and 200Δ10; Fig. 2A, c) show any association with Shc (Fig. 2A, a) or with the phosphatidylinositol-3'-OH kinase (PIk) 85K subunit (Fig. 2A, d). Similar conclusions were reached using a series of monoclonal antibodies that react with different regions of the MT molecule¹⁶. Antibodies that do not precipitate MT/pp60^{c-src} do

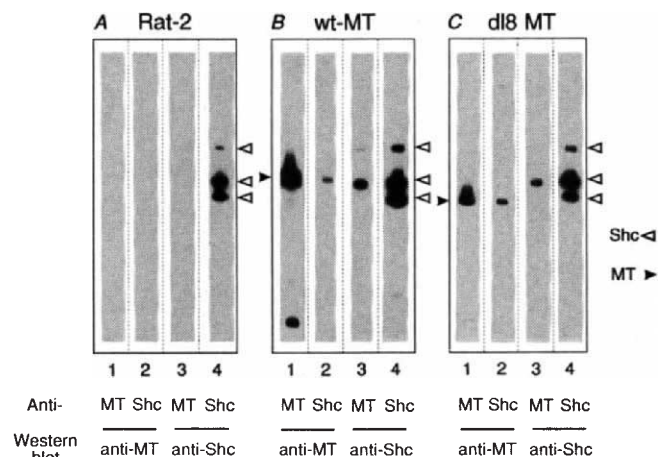


FIG. 1 Association of MT with Shc. Lysates of Rat-2 fibroblasts (A), wild-type (wt) MT transformed Rat-2 cells (B), or mutant dl8 MT-transformed Rat-2 cells (C), were immunoprecipitated with either an anti-MT monoclonal antibody or an anti-Shc polyclonal serum, western blotted and probed with either anti-MT or anti-Shc antibodies. The antibody combination used is indicated below each lane; migration positions are indicated for MT to the left (filled arrowhead) and for Shc 66K, 52K and 46K species to the right (arrowhead) of each panel.

METHODS. Confluent cells were lysed by the addition of 1 ml lysis buffer (100 mM Tris-Cl, pH 8.5, 100 mM NaCl, 1 mM Na₃VO₄, 0.5% Nonidet-P40 and 0.3 trypsin-inhibitor units (TIU) per ml aprotinin; Sigma) per 90-mm dish on ice for 30 min. 500 µl of each lysate was immunoprecipitated with either 50 µl of the new anti-MT monoclonal antibody PAb 762 (S.M.D., manuscript in preparation) containing tissue culture fluid or 2 µl of a polyclonal rabbit serum raised against the peptide APRDLDFMKPFEDALRVC, the sequence of amino acids 343–360 of Shc (ref. 26). Immune complexes were collected on protein A-Sepharose and washed three times in Tris-buffered saline (TBS: 150 mM NaCl, 50 mM Tris-Cl, pH 8.5, 0.05% Nonidet-P40, 0.3 TIU per ml aprotinin). Proteins were eluted by incubation with SDS sample buffer and separated by electrophoresis on an SDS-containing 10% polyacrylamide gel as described²⁷. The gel was then western-blotted onto Immobilon P membrane (Millipore) and duplicate strips cut from each filter. The remaining protein-binding sites were blocked with 5% dried skimmed milk in PBS. Strips were probed with biotinylated PAb 762 followed by horseradish peroxidase (HRP)-conjugated streptavidin (Jackson ImmunoResearch) to detect MT, or a rabbit anti-Shc SH2 domain serum⁹ followed by HRP-conjugated protein A to detect Shc. Development was with ECL reagent (Amersham International). Probing with an antibody having the same specificity as that used for precipitation provides a control to ensure even precipitation and to indicate the migration positions of MT and Shc (lanes 1 and 4).

not precipitate MT/Shc either (data not shown). Mutations in the C-terminal half of MT, however, have varying effects on each association. Mutants dl23, dl1015 and RX-2 bind Shc according to the amount of pp60^{c-src} bound (except possibly dl1015, which seems to bind less Shc); dl23 does not bind the PIK 85K subunit as the deletion removes Tyr 315 and the adjacent sequences of MT (Fig. 2A, d). This has no effect on Shc binding to MT, indicating that these two proteins bind to separate sites. Mutant 248M (ref. 15), however, has normal levels of pp60^{c-src} and PIK 85K subunit bound, but only interacts weakly with Shc. Together these data indicate that MT must bind pp60^{c-src} before associating with Shc and point to the region affected by mutant 248M as having a role in the MT/Shc interaction. As the 248M lesion is near a tyrosine residue phosphorylated by pp60^{c-src} (Tyr 250), we also investigated the properties of MT mutants having Tyr→Phe (Y→F) conversions at positions 250 or 315. These are defective in transforming ability, but are not completely negative^{17,18}. Both mutant MTs bind pp60^{c-src} (Fig. 2A, c), but the 250F mutant shows little association with Shc, although binding to mutant 315F is normal. Shc interaction, therefore, is probably with a region surrounding Tyr 250 of MT. This suggests that pp60^{c-src} is required to phosphorylate this residue of MT to provide a binding site for Shc, similar to the way in which PIK 85K is thought to bind Tyr 315 (refs 6–8). Mutations close to this region that retain transforming ability (dl8 and 210Δ10; our unpublished results) show no defect in Shc association. Although Shc binding to mutant dl1015 MT is reduced, it is not clear yet whether this is sufficient to account

for the transforming defect. Generally, however, Shc plus PIK 85K subunit binding shows a good correlation with the transforming ability of MT mutants, suggesting that both associations are required for transformation.

Shc proteins become phosphorylated on tyrosine residues during both *v-src* transformation¹² and activation of tyrosine kinase-associated growth-factor receptors (TKAGFR)⁹. Therefore the phosphorylation state of MT-associated Shc was examined. pp60^{c-src} in MT immunoprecipitates can phosphorylate MT-associated proteins^{19,21}. To determine whether this includes Shc, *in vitro*-labelled polypeptides were eluted and rebound with another antibody. Figure 3A, a (lane 1) shows the polypeptides phosphorylated in an anti-MT precipitate. The major species is MT itself, shown by its re-binding to another anti-MT antibody (lane 3). Re-precipitates formed with anti-Shc antibody contain a phosphorylated polypeptide resembling the 52K form of Shc (lane 5). As MT and the 52K form of Shc migrate similarly, the truncated MT encoded by mutant dl8 was again used to provide unambiguous identification (Fig. 3A, b). Similar phosphorylation patterns were observed, except the major species was truncated as expected. The polypeptide recognized by anti-Shc antibodies, however, remained unaltered, confirming it to be the 52K form of Shc. Re-precipitation with anti-phosphotyrosine antibodies shows that this polypeptide contains phosphotyrosine residues (lane 4). Shc, therefore, can be tyrosine-phosphorylated in MT precipitates.

To determine whether a similar phosphorylation occurs in cells, MT and Shc immunoprecipitates were blotted and probed

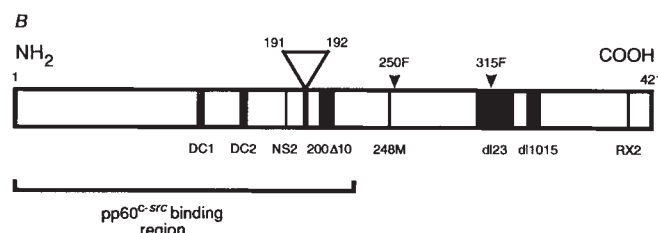
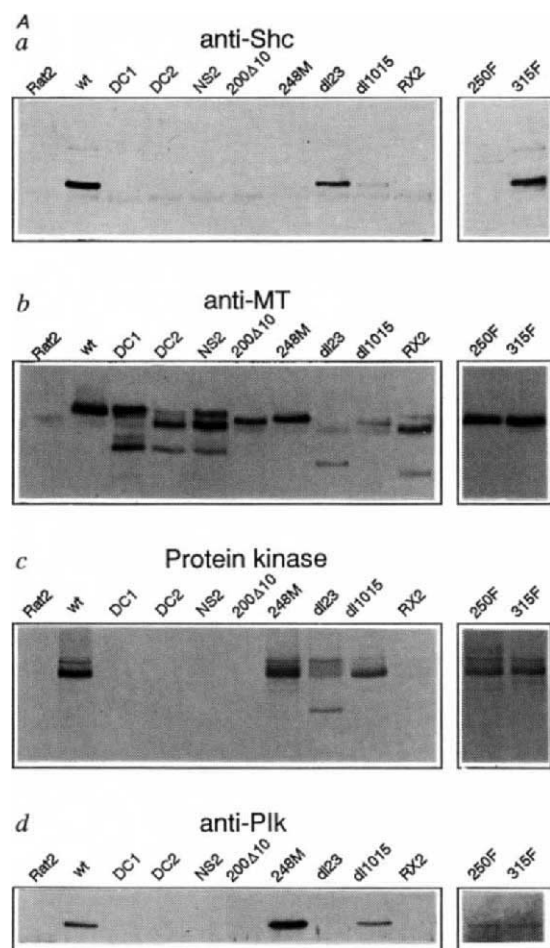


FIG. 2 Shc association with non-transforming mutants of MT. A, Lysates of Rat-2 cells expressing mutant forms of MT were immunoprecipitated with anti-MT PAb762, western blotted and probed with anti-Shc (a), anti-MT (b), or anti-PIK 85K subunit (d) antibodies. Duplicate immunoprecipitates were incubated with [γ -³²P]ATP before separation to detect pp60^{c-src} protein kinase activity (c). The mutant MT present in each lysate is indicated above each lane. B, Schematic representation of the MT sequences affected by the mutations used in A; see refs 14 and 15 for precise locations. Also marked is the binding region on MT for pp60^{c-src} defined by these mutants.

METHODS. 200Δ10 is a novel MT mutant with a deletion of the DNA encoding amino acids 200–209 of the sequence. It is non-transforming, binds PP2A, but not pp60^{c-src} (C.E.P.B., V. P. Horner and S.M.D., manuscript in preparation). Recombinant DNA clones expressing mutant MTs at residues 250 or 315 were constructed using the polymerase chain reaction (PCR) and mutant oligonucleotides essentially as in ref. 28. New mutant constructs were verified by DNA sequencing. Mutant 248M DNA was supplied by B. Druker; the remaining mutants were from W. Markland and A. Smith. Cell lines expressing mutant MT polypeptides were isolated by co-transfecting 10 μg MT plasmid (containing the *Bgl*I–*Eco*RI fragment of polyoma virus MT cDNA containing the appropriate mutation) together with 0.5 μg pSV2neo plasmid DNA per 90-mm dish of cells. Neo-expressing clones were selected by addition of G418 (GIBCO). Cell lines were single-cell cloned twice. 100 μl Rat-2 and wild-type MT and 1 ml of mutant-containing lysate (to correct for lower expression of mutant proteins) were immunoprecipitated with 25 μl monoclonal antibody containing tissue culture fluid as for Fig. 1. After western blotting, PVDF filters were probed with rabbit anti-Shc SH2 domain (A, a), biotinylated anti-MT PAb 762 (A, b), or rabbit anti-PIK 85K subunit (Transduction Labs; A, d) antisera, as for Fig. 1. Duplicate immunoprecipitates were incubated before elution with 2 μCi [γ -³²P]ATP (Amersham International) as in ref. 16.

with an anti-phosphotyrosine antibody. Figure 3B, a shows that in normal fibroblasts little phosphotyrosine is present on any of the polypeptides precipitated. When MT is expressed, however, a polypeptide migrating with MT in anti-MT precipitates is recognized (Fig. 3B, b, lane 2). Similarly, a polypeptide resembling Shc is detected in anti-Shc precipitates (lane 3). To confirm these identities, lysates of MT mutant dl8-expressing cells were used (Fig. 3B, c). Only two major phosphotyrosine-containing polypeptides are seen in both anti-MT and anti-Shc precipitates, one migrating with the truncated MT and the other with the 52K form of Shc. This shows that Shc is the principal tyrosine-phosphorylated species associated with MT in cells, again suggesting that this interaction has an important role in MT transformation, and that MT-associated Shc becomes tyrosine-phosphorylated *in vivo* as well as *in vitro*.

Shc phosphorylation by activated TKAGFRs results in formation of an Shc/Grb2 complex¹⁰. This may influence the activity of the mammalian Sos homologue, thus stimulating p21^{ras} activity (reviewed in ref. 11). As MT transformation is dependent upon Ras function^{22,23}, immunoprecipitates were western-blotted with anti-Grb2 antibodies to determine whether Shc phosphorylated by MT/pp60^{c-src} also binds Grb2. In untransformed cells, no association is detectable (Fig. 4A, lanes 1 and 3). However, Grb2 is present both in anti-MT and

anti-Shc precipitates from lysates of MT-transformed cells (lanes 2 and 4). The phosphorylation of Shc by the MT-bound pp60^{c-src} therefore promotes an interaction between Shc and Grb2.

Both Shc and Grb2 contain SH2 domains that can bind phosphotyrosine-containing polypeptides^{9,24}. To determine whether the interactions with MT involve these SH2 regions, polypeptides phosphorylated *in vitro* were incubated with fusion proteins containing the SH2 domains of Shc and Grb2 (Fig. 4B). Shc SH2 interacts with a protein resembling MT (Fig. 4B, a, lane 3), which is confirmed by reaction with truncated dl8 MT (Fig. 4B, b, lane 3). Grb2 SH2 also reacts with a phosphorylated MT-sized polypeptide (Fig. 4B, a, lane 4). However, this polypeptide does not alter in dl8 precipitates (Fig. 4B, b, lane 4) and so is probably Shc. The lack of reaction between MT and Grb2 supports the observation that the consensus recognition sequence of Grb2 SH2 (ref. 25) shows no homology with any site in MT. It is likely, therefore, that the MT/Shc/Grb2 complex involves, at least in part, the SH2 domain of Shc binding to phosphorylated tyrosine 250 of MT, and then the SH2 domain of Grb2 binding to tyrosine-phosphorylated Shc.

In conclusion, we have shown that in addition to binding and activating pp60^{c-src} and PI3K, MT also interacts with Shc. This probably involves the SH2 domain of Shc binding to Tyr 250 of

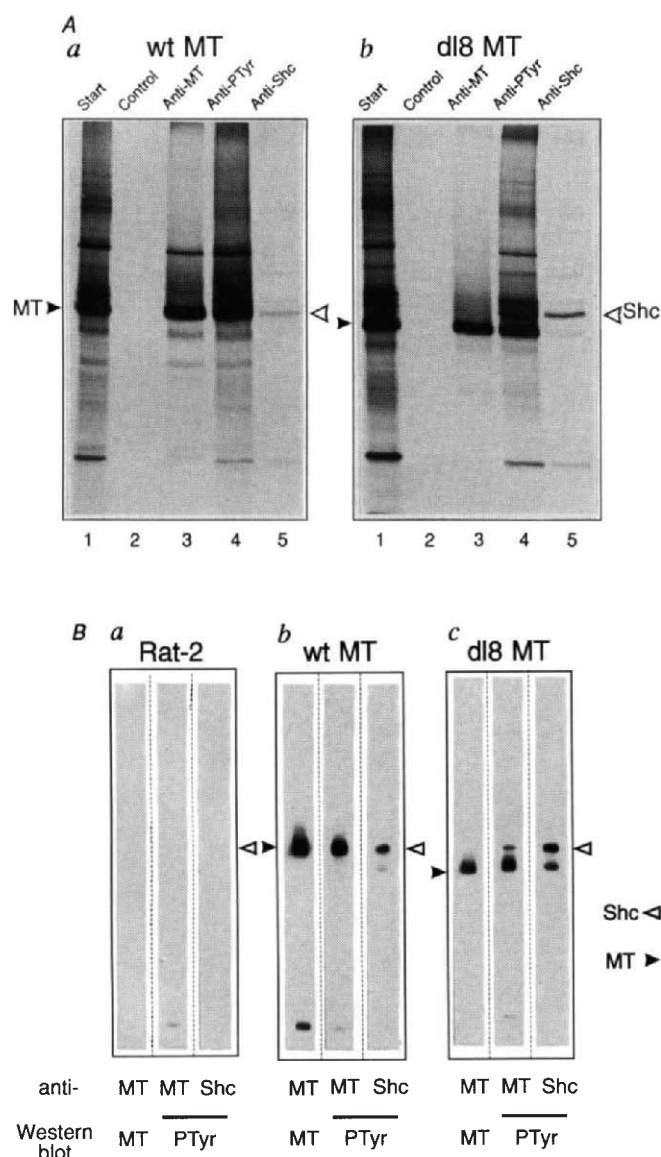
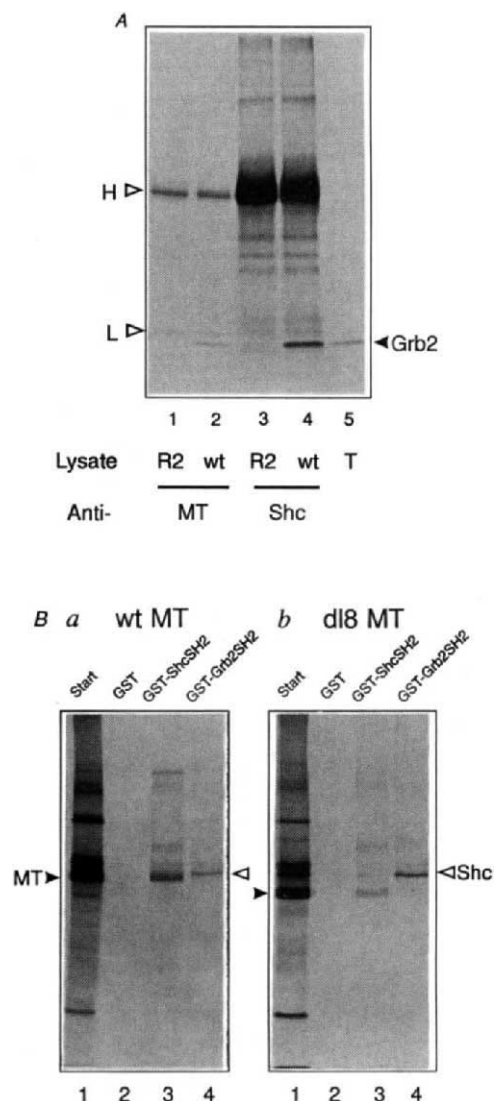


FIG. 3 Shc associated with MT becomes phosphorylated both *in vitro* and *in vivo*. **A**, MT immunoprecipitates formed with PAb 762 antibody and lysates from MT-transformed cells (**a**), or mutant dl8-transformed cells (**b**), were incubated with [γ -³²P]ATP, the phosphorylated polypeptides eluted, and then rebound with a second antibody. Lanes: 1, starting labelled polypeptides; 2, control reaction using no second antibody; 3, polypeptides re-precipitated with PAb 750 (ref. 16), or 4, with anti-phosphotyrosine monoclonal antibody PY20 (Transduction Labs), or 5, with anti-Shc peptide serum. Migration positions of MT and the 52K form of Shc are indicated. **B**, Lysates of Rat-2 (**a**), MT-transformed (**b**), or mutant dl8-transformed (**c**) cells were immunoprecipitated with anti-MT or anti-Shc antibodies, western blotted and then probed with recombinant anti-phosphotyrosine antibody RC20 (Transduction Labs). The antibodies used for precipitation and blotting are indicated under each lane. A control reaction using MT immunoprecipitates probed with anti-MT is shown to indicate the migration position of MT (lanes 1). The migration positions of MT and the 52K form of Shc are indicated.

METHODS. **A**, 500 μ l MT- (**a**) or mutant dl8- (**b**) containing cell lysate were immunoprecipitated with 50 μ l PAb 762 containing tissue culture fluid as for Fig. 1. The resulting precipitates were incubated with 25 μ Ci [γ -³²P]ATP as for Fig. 2. Polypeptides were then eluted by incubation for 5 min at room temperature with 100 μ l 1% SDS, 80 mM Tris-Cl, pH 6.8. Supernatants were removed, boiled for 5 min, and divided into five aliquots. Sample **a** was untreated; samples **b**–**e** were diluted with 500 μ l TBS buffer containing 0.5% Nonidet-P40 and chilled to 4 °C. Renatured proteins were then re-precipitated without the addition of fresh antibody (lanes 2), with 30 μ l PAb 750 TCF (lanes 3), 2 μ g anti-phosphotyrosine monoclonal antibody PY20 (lanes 4; Transduction Labs), or 0.5 μ l Shc anti-peptide antiserum (lanes 5). Immune complexes were collected on protein A-Sepharose, washed with TBS, eluted and then electrophoresed as for Fig. 1. The gel was dried and autoradiographed. **B**, 500 μ l cell lysate was immunoprecipitated and blotted as for Fig. 1. Strips were cut from the filters, blocked and probed with recombinant antibody RC20 conjugated to horseradish peroxidase according to the manufacturer's instructions (Transduction Labs); development was with ECL reagent.

FIG. 4 Shc can bind to MT and Grb2 to Shc through SH2 domains. **A**, Polypeptides were immunoprecipitated from lysates of Rat-2 (R2) or MT-transformed Rat-2 cells (wt) with anti-MT or anti-Shc antibodies, western blotted and probed with anti-Grb2 antiserum. The lysate and antibody used for precipitation are indicated below each lane. The migration position of Grb2 is indicated on the right. Other bands are due to reaction with the heavy and light chains (H and L) of the antibodies used for the immunoprecipitation (left). Total starting lysate (T; lane 5) was also loaded as a migration control. **B**, MT immunoprecipitates were incubated with [γ - 32 P]ATP and the polypeptides eluted as for Fig. 3B. The renatured proteins were then bound to the glutathione S-transferase (GST) fusion protein indicated above each lane. Lanes: 1, starting labelled polypeptides; 2, proteins bound to GST alone; 3, proteins bound to GST-ShcSH2; and 4, proteins bound to GST-Grb2SH2 fusion proteins. Migration positions of MT and the 52K form of Shc are indicated. **METHODS.** **A**, 100 μ l of Rat-2 or wild-type MT-transformed cell lysates were immunoprecipitated with PAb 762 anti-MT monoclonal antibody or an anti-Shc C-terminal peptide antiserum and western blotted as for Fig. 1. In addition, 1 μ l total Rat-2 cell lysate was loaded (lane 5). The filter was then probed with antiserum raised against a Grb2-GST fusion protein, alkaline phosphatase-conjugated goat anti-rabbit antibodies (Sigma), and developed with bromochloroindolyl/nitroblue tetrazolium by standard methods. **B**, Immunoprecipitated polypeptides were labelled, eluted and renatured as in **A**. Eluates were left untreated (lane 1), or reacted with 10 μ g GST fusion proteins bound to glutathione-Sepharose (Pharmacia). The cDNAs corresponding to the Shc and Grb2 SH2 domains were isolated by PCR and cloned into pGEX-2T (ref. 29). Bacterial lysates containing the expressed fusion protein were isolated essentially as in ref. 29. Clarified bacterial lysates were incubated with glutathione-Sepharose for 30 min on ice, the Sepharose isolated and washed in TBS. After incubation for 1 h at 4 $^{\circ}$ C with the renatured labelled polypeptides, the Sepharose was isolated, washed, and the bound proteins eluted, electrophoresed and autoradiographed as for Fig. 3A.



MT which has been phosphorylated by pp60^{c-src}. Shc is thus brought close to pp60^{c-src} and becomes tyrosine-phosphorylated itself. This provides a binding site for Grb2, which presumably stimulates p21^{ras} activity through the mammalian Sos similar to activated TKAGFRs. Repositioning the Sos homologue to the membrane, and therefore close to p21^{ras}, has been proposed as a mechanism for p21^{ras} activation¹¹. MT is also located on the membrane, so could act similarly. MT and activated TKAGFRs

therefore have several properties in common. Both are membrane-bound and stimulate a tyrosine kinase activity that phosphorylates multiple tyrosine sites in the molecule. These phosphorytyrosines act as binding sites for SH2-containing proteins, which in turn are tyrosine-phosphorylated. The similarities are so striking that MT could be considered as a permanently activated TKAGFR and so present MT as a system for studying signal transduction pathways. □

Received 15 September; accepted 29 November 1993.

- Treisman, R., Novak, U., Favaloro, J. & Kamen, R. *Nature* **292**, 595-600 (1981).
- Courtneidge, S. A. *Cancer Surv.* **5**, 173-182 (1986).
- Kaplan, D. R., Pallas, D. C., Morgan, W., Schaffhausen, B. & Roberts, T. M. *Biochim. biophys. Acta* **948**, 345-364 (1989).
- Harvey, R., Oostra, B. A., Belsham, G. J., Gillett, P. & Smith, A. E. *Molec. cell. Biol.* **4**, 1334-1342 (1984).
- Hunter, T., Hutchinson, M. A. & Eckhart, W. *EMBO J.* **3**, 73-79 (1984).
- Talmage, D. A. et al. *Cell* **59**, 55-65 (1989).
- Yokim, M. et al. *J. Virol.* **66**, 5485-5491 (1992).
- Auger, K. R., Carpenter, C. L., Shoelson, S. E., Piwnicka-Worms, H. & Cantley, L. C. *J. Biol. Chem.* **267**, 5408-5415 (1992).
- Pellicci, G. et al. *Cell* **70**, 93-104 (1992).
- Rozakis-Adcock, M. et al. *Nature* **360**, 689-692 (1992).
- McCormick, F. *Nature* **363**, 15-16 (1993).
- McGlade, J., Cheng, A., Pellicci, G., Pellicci, P. G. & Pawson, T. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8869-8873 (1992).
- Griffin, B. E. et al. *Cold Spring Harb. Symp. quant. Biol.* **1**, 271-283 (1980).
- Markland, W. & Smith, A. E. *Biochim. biophys. Acta* **907**, 299-321 (1987).
- Druker, B. J., Ling, L. E., Cohen, B., Roberts, T. M. & Schaffhausen, B. S. *J. Virol.* **64**, 4454-4461 (1990).

- Dilworth, S. M. & Horner, V. P. *J. Virol.* **67**, 2235-2244 (1993).
- Carmichael, G., Schaffhausen, B. S., Mandel, G., Liang, T. J. & Benjamin, T. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 679-683 (1984).
- Markland, W. et al. *J. Virol.* **59**, 384-391 (1986).
- Dilworth, S. M. *EMBO J.* **1**, 1319-1328 (1982).
- Courtneidge, S. A. & Heber, A. *Cell* **50**, 1031-1037 (1987).
- Kaplan, D. R. et al. *Cell* **50**, 1021-1029 (1987).
- Raptis, L. et al. *J. Virol.* **65**, 5203-5210 (1991).
- Jelinek, M. A. & Hassell, J. A. *Oncogene* **7**, 1687-1698 (1992).
- Lowenstein, E. J. et al. *Cell* **70**, 431-442 (1992).
- Songyang, Z. et al. *Cell* **72**, 767-778 (1993).
- Segatto, O. et al. *Oncogene* **8**, 2105-2112 (1993).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. *Gene* **77**, 51-59 (1989).
- Smith, D. B. & Johnson, K. S. *Gene* **67**, 31-40 (1988).

ACKNOWLEDGEMENTS. We thank W. Markland and A. E. Smith for polyoma virus mutant DNA clones, and B. J. Druker for mutant 248M DNA. This work was supported by the Cancer Research Campaign (UK).

DNA repair gene *RAD3* of *S. cerevisiae* is essential for transcription by RNA polymerase II

Sami N. Guzder^{*†}, Hongfang Qiu^{*†},
Christopher H. Sommers^{*†}, Patrick Sung^{*†},
Louise Prakash^{†‡} & Satya Prakash^{*†§}

^{*} Department of Biology and [†] Department of Biophysics, University of Rochester, Rochester, New York 14642, and [‡] Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, Texas 77555, USA

[§] Present addresses: Sealy Center for Molecular Science (S.N.G., H.Q., P.S., L.P. and S.P.); and Xenometrix Inc., Boulder, Colorado 803301, USA (C.H.S.).

[§] To whom correspondence should be addressed at University of Texas.

THE *RAD3* gene of *Saccharomyces cerevisiae* is required for excision repair of ultraviolet-damaged DNA and is essential for cell viability¹. The *RAD3*-encoded protein shares a high degree of homology with the human *ERCC2(XPD)* gene product². Mutations in *XPD*, besides causing the cancer-prone syndrome xeroderma pigmentosum, can also result in Cockayne's syndrome and trichothiodystrophy³. To investigate the role of *RAD3* in viability, we examine here the effect of a recessive, temperature-sensitive (*ts*) conditional lethal mutation of the gene on transcrip-

tion by RNA polymerase II. Upon transfer to the restrictive temperature, the *rad3-ts* mutant rapidly ceases growth and poly(A)⁺ RNA synthesis is inhibited drastically. Messenger RNA levels of all the genes examined, *HIS3*, *TRP3*, *STE2*, *MET19*, *RAD23*, *CDC7*, *CDC9* and *ACT1*, decline rapidly upon loss of *RAD3* activity. The synthesis of heat-shock-inducible *HSP26* mRNA and galactose-inducible *GAL7* and *GAL10* mRNAs is also drastically inhibited in the *rad3-ts* mutant at the restrictive temperature. The RNA polymerase II transcriptional activity in extract from the *rad3-ts₁₄* strain is thermolabile, and this *in vitro* transcriptional defect can be fully corrected by the addition of homogeneous *RAD3* protein. These findings indicate that *RAD3* protein has a direct and essential role in RNA polymerase II transcription.

The *rad3-ts₁₄* mutant strain grows at the permissive temperature (25 °C), but fails to grow at the non-permissive temperature (39 °C) (Fig. 1a). Introduction of the wild-type *RAD3* gene carried on the low-copy-number plasmid pDG255 into the *rad3-ts₁₄* strain restored normal growth at 39 °C (Fig. 1a). Upon shift to 39 °C, the mutant strain rapidly ceases growth, with cells undergoing less than one division (Fig. 1b). By appropriate subcloning, we mapped the *rad3-ts₁₄* mutation to a 410-base-pair (bp) *Bcl*II/*Eco*RI fragment encompassing nucleotides +1,510 to +1,919 (ref. 4). Sequence analysis of this DNA fragment revealed a single mutation in codon 590, changing it from TGC to TAC, resulting in the substitution of cysteine by tyrosine.

The *rad3-ts₁₄* mutation affects the synthesis of total RNA, as determined by pulse-labelling of cells with [³H]uracil before and after shift to 39 °C (Fig. 1c). To determine whether the *rad3-ts₁₄* mutation affects transcription by RNA polymerase II, poly(A)⁺ RNA was purified from total RNA pulse-labelled with

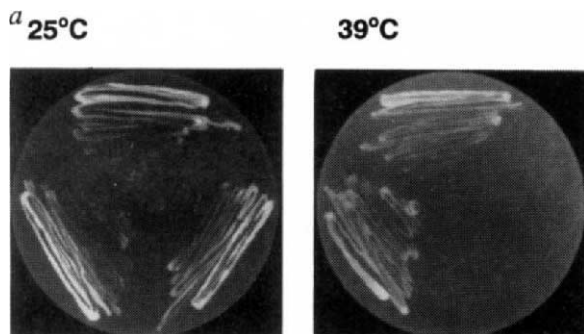
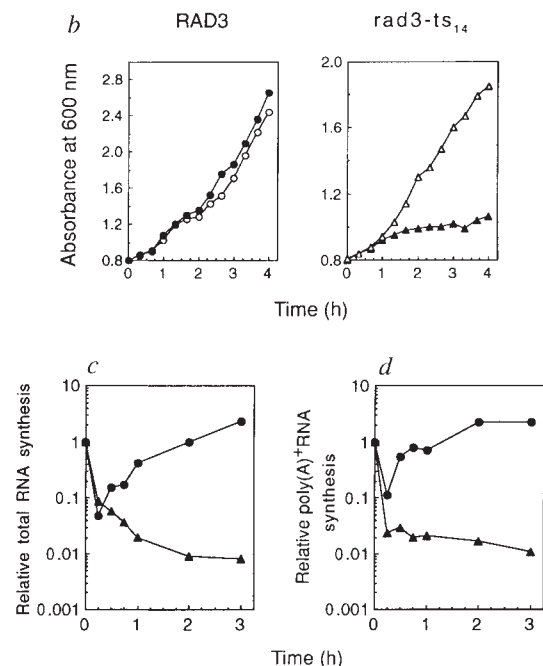


FIG. 1 Effect of the *rad3-ts₁₄* mutation on growth, total RNA synthesis and poly(A)⁺ RNA synthesis. **a**, Growth of wild type and *rad3-ts₁₄* mutant strains on YPD plates at 25 °C and 39 °C. Top, wild-type strain RSY11 (*RAD3*); lower right, *rad3-ts₁₄* mutant strain; lower left, *rad3-ts₁₄* strain carrying the *RAD3* gene on plasmid pDG255. **b**, Growth of *RAD3* and *rad3-ts₁₄* strains in synthetic complete (SC) medium as determined by measuring the A_{600} at various times. Symbols: *RAD3* strain at 25 °C (○) and at 39 °C (●); *rad3-ts₁₄* strain at 25 °C (△) and at 39 °C (▲). **c**, Total RNA synthesis in *RAD3* and *rad3-ts₁₄* strains. Total cellular RNA was isolated from each [³H]uracil pulse-labelled cell sample. Radioactivity in the same amount of total RNA (2 µg) of each sample was determined by liquid scintillation counting. For the samples pulse-labelled at 39 °C, the incorporation of [³H]uracil (c.p.m. per µg total RNA) was normalized by the counts present in the sample that was pulse-labelled at 25 °C (time 0). Symbols: *RAD3*, ●; *rad3-ts₁₄*, ▲. **d**, poly(A)⁺ RNA synthesis. Poly(A)⁺ RNA was isolated from the same amount of [³H]uracil pulse-labelled total RNA (100 µg) by PolyAtract mRNA purification system (Promega). Radioactivity in isolated poly(A)⁺ RNA was determined by liquid scintillation counting. For the samples pulse-labelled at 39 °C, incorporation of [³H]uracil was normalized by the counts in the sample that was pulse-labelled at 25 °C (time 0). Symbols: *RAD3*, ●; *rad3-ts₁₄*, ▲.

METHODS. **a** and **b**, The media used were YPD and synthetic complete (SC). The *rad3-ts₁₄* mutant allele was isolated using hydroxylamine mutagenesis and the plasmid shuffle technique (R. Schiestl and S.P., unpublished). The *rad3-ts₁₄* mutation was mapped by subcloning and its nucleotide sequence determined. The strains used here are RSY11



(*RAD3*) and its temperature-sensitive mutant derivative RSY11 (*rad3-ts₁₄*), in which the mutant *rad3-ts₁₄* gene has been integrated into the chromosome, replacing the wild-type chromosomal copy of the *RAD3* gene. **c** and **d**, *RAD3* and *rad3-ts₁₄* strains were grown at 25 °C in SC medium to $A_{600} \sim 0.6$ and divided into several aliquots. [5,6-³H]uracil was added to one aliquot and cells pulse-labelled at 25 °C for 10 min ('0 time' sample). The other aliquots were incubated in prewarmed flasks at 39 °C. The 10-min pulse-labelling was at 39 °C at 15, 30, 45, 60, 120 and 180 min after the shift to 39 °C. The final concentration of [5,6-³H]uracil used for pulse-labelling was 30 µCi ml⁻¹. After pulse-labelling, cultures were quickly chilled by pouring over frozen crushed 1M sorbitol. Cells were subsequently pelleted, quickly frozen in crushed ice, and used for preparing total ³H-labelled RNA as described in the legend to Fig. 2.

[³H]uracil. As shown in Fig. 1d, poly(A)⁺ RNA synthesis was drastically inhibited at the restrictive temperature in the *rad3-ts₁₄* strain.

The severe reduction in poly(A)⁺ RNA synthesis in the *rad3-ts₁₄* mutant at 39 °C indicated that the mutant cells are defective in mRNA synthesis. To investigate further the effect of the *rad3-ts₁₄* mutation in RNA polymerase II transcription, we used northern blot analysis to examine the levels of *HIS3*, *TRP3*, *STE2*, *MET19*, *RAD23*, *CDC7*, *CDC9* and *ACT1* mRNAs in the wild-type and *rad3-ts₁₄* strain before (0 min sample) and after shifting the cultures to 39 °C (Fig. 2a). These genes affect diverse cellular processes such as amino-acid biosynthesis, DNA repair, DNA replication and cytoskeleton formation. As the steady-state levels of mRNAs are a function of rates of their synthesis and degradation, a defect in mRNA synthesis in the *rad3-ts₁₄* mutant would cause a reduction in the level of mRNAs. In the absence of new mRNA synthesis, the rate of depletion of mRNAs at the restrictive temperature would be a function of their half-lives. Transfer of strains carrying the *rpb1-1* mutation in the gene encoding the largest subunit of RNA polymerase II, or carrying *ts* mutations in the gene encoding the TATA-binding protein (TBP), to the restrictive temperature results in a decline in the steady-state levels of mRNAs^{5,6}. We included genes that encode mRNAs with short or relatively long half-lives; for example, *HIS3* mRNA has a half-life of ~7 min, and *ACT1* mRNA has a half-life of ~25 min (ref. 7). Transfer of the *rad3-ts₁₄* strain to 39 °C resulted in a drastic reduction in the level of *HIS3* transcripts. Levels of *TRP3* and *STE2* mRNAs were also greatly diminished upon shift of the *rad3-ts₁₄* strain to the restrictive temperature. In wild-type cells, *MET19* mRNA levels showed no decline upon temperature shift and the levels remained constant during the entire 3-hour period at 39 °C; in

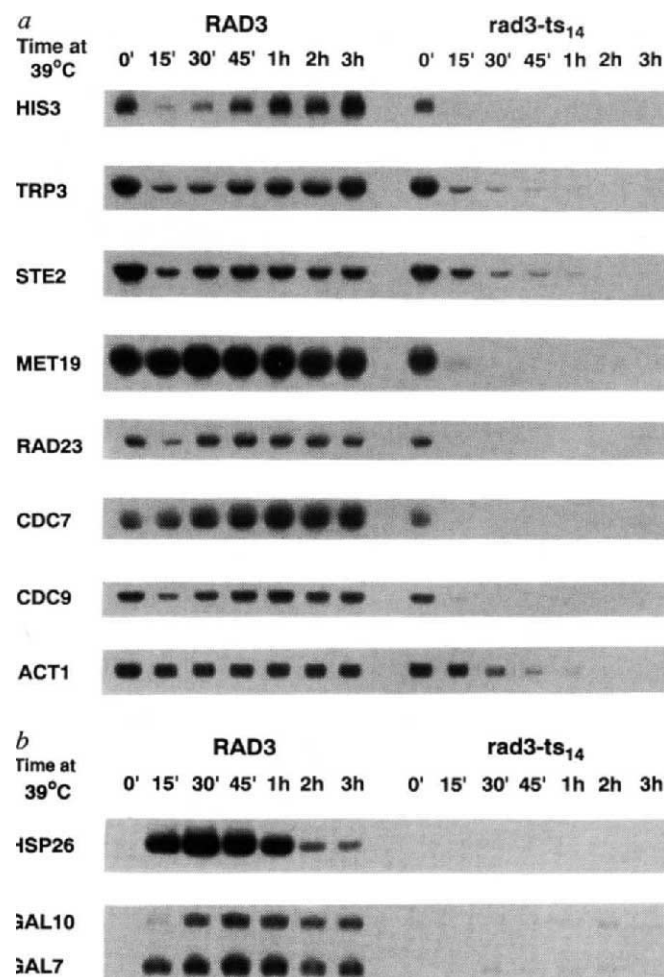
contrast, in the *rad3-ts₁₄* strain, *MET19* mRNA levels fell rapidly upon transfer of the culture to 39 °C. mRNA levels of the DNA repair gene *RAD23* also fell sharply upon transfer of the *rad3-ts* mutant to 39 °C. *CDC7* encodes a serine (threonine) protein kinase required for initiation of DNA replication. Upon shift to 39 °C, the *rad3-ts₁₄* strain shows a rapid and severe reduction in the level of *CDC7* mRNA, whereas no reduction is observed in the wild-type strain. The mRNA levels of the DNA ligase gene *CDC9* are also severely affected by transfer of the *rad3-ts₁₄* culture to 39 °C. In addition, we examined the levels of *ACT1*-encoded mRNA in the wild-type and *rad3-ts₁₄* strains. Whereas the wild-type strain had constant levels of *ACT1* mRNA throughout the 3-h incubation at 39 °C, *ACT1* mRNA levels fell gradually in the *rad3-ts₁₄* strain and were undetectable by 2 h.

We next determined the effect of the *rad3-ts₁₄* mutation on inducible synthesis of *HSP26*, *GAL7* and *GAL10* mRNAs (Fig. 2b). Transfer of wild-type culture from 25 °C to 39 °C resulted in a dramatic increase in heat-shock-inducible *HSP26* mRNA, with peak levels being attained by 30 min. In contrast, *HSP26* mRNA was not induced in the *rad3-ts₁₄* mutant. To test the induction of *GAL7* and *GAL10* mRNAs, galactose was added and cultures transferred immediately to 39 °C. The synthesis of *GAL7* and *GAL10* mRNAs was severely hampered by the *rad3-ts₁₄* mutation. These observations indicate that activated transcription mediated by RNA polymerase II is also greatly inhibited in the *rad3-ts₁₄* mutant at the restrictive temperature. A general requirement for *RAD3* in RNA polymerase II transcription *in vivo* is therefore implicated.

To determine the effect of the *rad3-ts₁₄* mutation on RNA polymerase III transcription, we examined the transcription of the yeast isoleucine transfer RNA gene. To circumvent the high

FIG. 2 Effect of the *rad3-ts₁₄* mutation on mRNA levels of various yeast genes. a, mRNA levels of *HIS3*, *TRP3*, *STE2*, *MET19*, *RAD23*, *CDC7*, *CDC9* and *ACT1* genes in the *RAD3* and *rad3-ts₁₄* strains. Cultures were grown at 25 °C in SC medium, transferred to 39 °C, and mRNA levels determined at the indicated times by northern hybridization. b, Heat-shock-inducible synthesis of *HSP26* mRNA and galactose-inducible synthesis of *GAL7* and *GAL10* mRNAs is inhibited at the restrictive temperature in the *rad3-ts₁₄* mutant strain. Cultures were grown at 25 °C in SC medium, transferred to 39 °C, and *HSP26* mRNA quantified after transfer to 39 °C. For *GAL7* and *GAL10* mRNA determination, cultures were grown at 25 °C in YP5% glycerol medium containing 1% yeast extract, 2% peptone and 5% (v/v) glycerol. Galactose was added to a final concentration of 2% and cultures were immediately transferred to 39 °C.

METHODS. Total cellular RNA was prepared as described²⁹. Briefly, cell pellets were resuspended in 350 µl yeast lysis buffer (200 mM Tris-HCl, pH 7.5, 500 mM NaCl, 10 mM EDTA, 0.5% SDS, 10 mM iodoacetic acid and 10 mM vanadyl ribonucleoside complex (BRL), and broken by glass beads with 1 vol phenol/chloroform (1/1). RNA was precipitated by ethanol, treated with RNase-free DNase I (Boehringer), and quantified by spectrophotometry. An equal amount of RNA from each sample was denatured by glyoxal, fractionated in a 1.4% agarose gel and electroblotted onto GeneScreen membrane (Dupont, NEN). DNA fragments used as hybridization probes for detecting various mRNAs were ³²P-labelled using the multiprime DNA-labelling system (Amersham). mRNA in autoradiograms was quantified by densitometry.



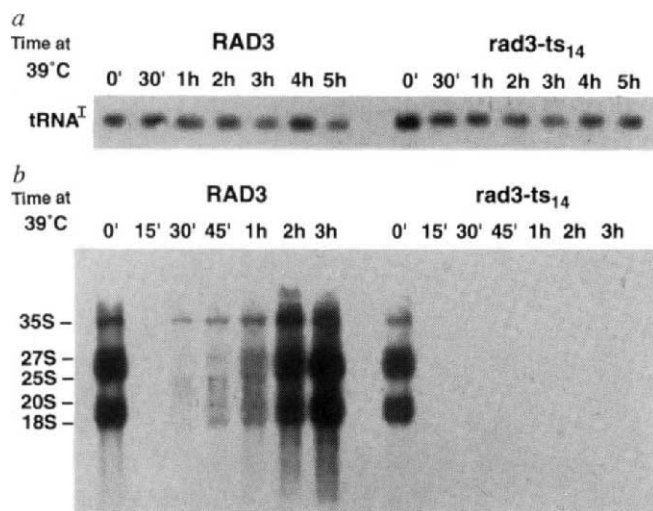


FIG. 3 Effect of the *rad3-ts₁₄* mutation on transcription by RNA polymerase III and RNA polymerase I. *a*, Levels of intron-containing precursor of isoleucine tRNA (*tRNA^{Ile}*) transcribed by RNA polymerase III in the *RAD3* and *rad3-ts₁₄* strains after shifting to 39 °C. *b*, Levels of rRNAs transcribed by RNA polymerase I in the *RAD3* and *rad3-ts₁₄* strains after shift to 39 °C.

METHODS. *a*, To examine transcription of the *tRNA^{Ile}* gene, for each sample 2 µg total RNA, prepared as described for Fig. 2, was fractionated in a 3% NuSieve 3:1 agarose gel (FMC), and the fractionated RNA was electrophoretically transferred onto GeneScreen. Intron-containing precursor *tRNA^{Ile}* was detected by northern hybridization to the ³²P-labelled oligonucleotide 5'-AAAGGCCTGTTTGAAGGCTTTGGC-ACAGAACTTCGGAACCG-3' (ref. 30) which contains only the internal portion of the intron. The hybridization probe was in excess, as verified by decreasing the probe >5-fold without affecting the band intensities. *b*, Total RNA (20 µg) from cells pulse-labelled with [5,6-³H]uracil (Fig. 1 legend) was electrophoresed in a 3% NuSieve 3:1 agarose gel. The gel was treated with Amplify (Amersham), dried and fluorographed at -70 °C with an intensifying screen.

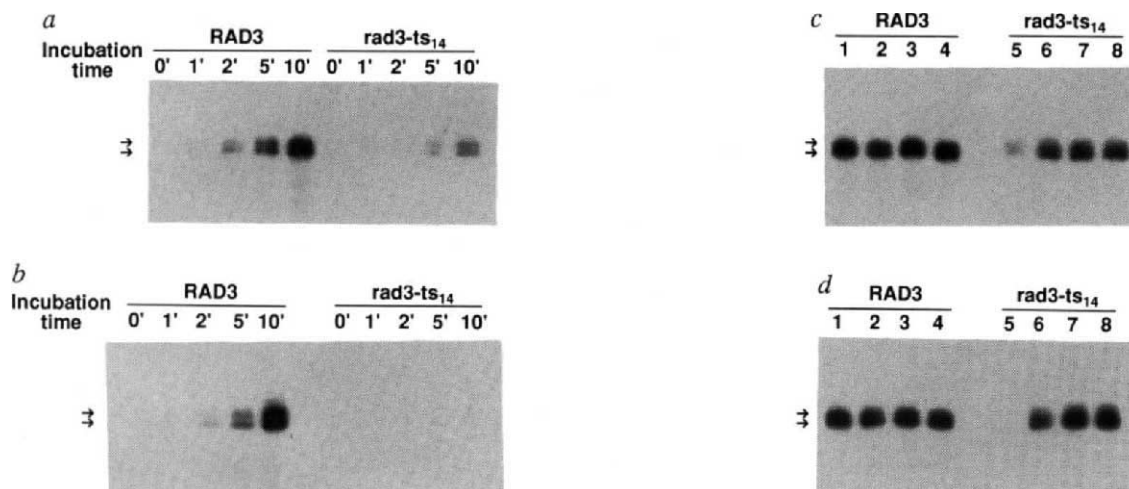


FIG. 4 *RAD3* protein is required for RNA polymerase II transcription *in vitro*. *a*, Time-dependent RNA polymerase II transcription in extracts of isogenic wild-type *RAD3* and *rad3-ts₁₄* yeast strains. Reactions were done at 22 °C for the times indicated. *b*, The residual capacity for pol II transcription in *rad3-ts₁₄* extract is thermolabile. Wild-type and *rad3-ts₁₄* extracts were preincubated at 39 °C for 5 min before being assayed for pol II transcription at 22 °C for the indicated times. *c* and *d*, Purified *RAD3* protein restores pol II transcriptional activity in *rad3-ts₁₄* extract. Wild-type and *rad3-ts₁₄* extracts without (*c*) or with (*d*) a 5-min pretreatment at 39 °C were tested for pol II transcription at 22 °C for 10 min in the absence of added *RAD3* (lanes 1 and 5) or in the presence of 5 ng (lanes 2 and 6), 10 ng (lanes 3 and 7) and 20 ng (lanes 4 and 8) of homogeneous *RAD3* protein. Arrows indicate the 375- and 350-nucleotide transcripts¹⁰.

METHODS. *RAD3* protein was purified to homogeneity from a yeast strain overproducing the protein as described¹¹. Extracts highly competent for pol II transcription were prepared from yeast cells grown at 25 °C and collected during the logarithmic growth phase according to ref. 9. Protein concentrations in extracts were determined with BCA reagent

(Pierce) using BSA as standard and varied between 55 and 75 mg ml⁻¹. Transcription was assayed as described⁹ using the DNA template pSL187, which contains a 400-bp G-less region downstream of the yeast *CYC1* TATA element and yields transcripts of 375 and 350 nucleotides¹⁰. Extract (200 µg protein) was preincubated with template DNA (1 µg) in reaction buffer (50 mM HEPES, pH 7.5, 10% glycerol, 90 mM potassium glutamate, 10 mM magnesium acetate, 0.75% w/v polyethylene glycol, *M_r* 3,350, 5 mM EGTA, 2.5 mM DTT, 1.4 U ml⁻¹ creatine kinase, 30 mM creatine phosphate and 40 U RNasin (Promega) in 25 µl for 5 min at 22 °C. Transcription was initiated by addition of 5 µl nucleotide mix containing 2.5 mM ATP, 2.5 mM CTP, 25 µM UTP and 10 µCi [³²P]UTP (800 Ci mmol⁻¹; Amersham). After incubation at 22 °C, reaction mixtures were treated with 6 U RNase T1, deproteinized, and the transcripts precipitated overnight at -70 °C with 3 vol ethanol with 10 µg yeast tRNA as carrier⁹. Transcripts were dissolved in 30 µl loading buffer (80% formamide/0.1 × TBE/0.01% xylene cyanol FF/0.01% bromophenol blue) and 10 µl electrophoresed in a 4% polyacrylamide gel containing 7 M urea. The ³²P-labelled transcripts were visualized by autoradiography of dried gels.

stability problem of tRNAs, we used a hybridization probe complementary to the intron sequence present in the *tRNA^{Ile}* gene. As tRNA introns are processed very rapidly (half-life <3 min)^{6,8}, the level of precursor tRNA containing the intron sequence should reflect the rate of transcription of the gene. As shown in Fig. 3*a*, the levels of precursor *tRNA^{Ile}* were not affected by the *rad3-ts₁₄* mutation when cells were incubated at the restrictive temperature.

The effect of the *rad3-ts₁₄* mutation on the synthesis and accumulation of RNA polymerase I-dependent ribosomal RNAs was assessed by pulse-labelling total yeast RNA with [³H]uracil and subjecting the RNA to gel electrophoresis and fluorography. The results in Fig. 3*b* indicate that the synthesis of 35S rRNA precursor and the accumulation of fully processed 25S and 18S rRNAs was greatly diminished in the mutant after a shift to 39 °C.

In its effects on transcription by RNA polymerases I, II and III, the *rad3-ts₁₄* mutation resembles the *rpb1-1* mutation of the gene encoding the largest subunit of polymerase II. Using methodology similar to ours, Cormack and Struhl found no effect of the *rpb1-1* mutation on RNA polymerase III transcription⁶. In contrast, *ts* mutations in the *TBP* gene cause a marked decrease in the transcription of yeast tRNA genes⁶. Both the synthesis of 35S rRNA and accumulation of mature rRNAs are greatly affected by the *rpb1-1* mutation^{5,6}. The shut-down of rRNA synthesis in the *rpb1-1* mutant may be a result of a stringent response to amino-acid deprivation⁵. The reduction in rRNA synthesis in the *rad3-ts₁₄* mutant could also be

due to the stringent response rather than to a direct involvement of RAD3 in transcription by RNA polymerase I, but this last possibility cannot be entirely ruled out.

To determine whether RAD3-encoded protein is directly involved in RNA polymerase II transcription, we prepared extracts⁹ from isogenic wild-type RAD3 and *rad3-ts₁₄* strains grown at 25 °C and tested their ability to carry out pol II transcription⁹ using the template pSL187 which contains the yeast *CYC1* TATA element and a 400-bp G-less region¹⁰. Extracts from the *rad3-ts₁₄* strain carry out RNA polymerase II transcription at a reduced rate compared to wild-type extracts (Fig. 4a). Exposure of extracts for 5 min to 39 °C does not affect the transcriptional ability of the wild-type RAD3 extract significantly, but this brief heat treatment abolishes the transcriptional competence of the *rad3-ts₁₄* extract (Fig. 4b). These observations indicate that RNA polymerase II transcription in *rad3-ts₁₄* extract is thermolabile. Addition of 5 ng homogeneous RAD3 protein¹¹ does not affect RNA polymerase II transcription in wild-type extracts, regardless of whether they are pretreated at 39 °C for 5 min (Fig. 4c, d). Strikingly, addition of 5 ng RAD3 protein to *rad3-ts₁₄* extracts that have or have not been exposed to 39 °C for 5 min restores the ability of mutant extracts to carry out RNA polymerase II transcription to wild-type levels (Fig. 4c, d). These observations indicate that RAD3 functions directly in RNA polymerase II transcription.

Initiation of transcription by RNA polymerase II requires the binding of the transcription factor TBP (TFIID) to the TATA element followed by the binding of transcription factors TFIIA, TFIIB, TFIIE, TFIIIF, TFIIJ and TFIIH at the promoter site^{12,13}. Conversion of this preinitiation complex to an active form capable of initiating transcription requires the hydrolysis of the β - γ bond of ATP^{14,15}. In addition to promoting phosphorylation of the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II, ATP hydrolysis serves another function, because basal transcription occurs in the absence of CTD phosphorylation¹⁶. In this other role, ATP hydrolysis could be required for melting the DNA to form the open complex¹⁷. Transcription factor *b* from yeast, factor δ from rat liver and factor TFIIH (BTF2) from humans resemble one another and they all have a DNA-dependent ATPase activity^{18–20}. The *XPB(ERCC3)* gene product is associated with TFIIH and a DNA helicase activity copurifies with this transcription factor²⁰. The *RAD25* gene of *S. cerevisiae* encodes the yeast counterpart of XPB protein and both proteins contain the conserved helicase sequence motifs consistent with the possibility of a helicase activity in these proteins^{21,22}. The RAD3 protein possesses DNA-dependent ATPase, DNA helicase and DNA–RNA helicase activities^{11,23,24}. The *RAD25* and *RAD3* genes further resemble one another in that besides their requirement in nucleotide excision repair, both are essential for cell viability^{1,21,22}. The RAD3 and RAD25 proteins may function together in unwinding the duplex DNA at the site of transcription initiation by RNA polymerase II.

The *XPD* gene is the human counterpart of the yeast *RAD3* gene. In addition to causing xeroderma pigmentosum, mutations in the *XPD* gene can also result in Cockayne's syndrome and trichothiodystrophy³. The severe dwarfism, mental retardation, and cachexia in Cockayne's syndrome patients and sulphur-deficient brittle hair in trichothiodystrophy patients²⁵ could arise from deficiencies in transcription resulting from mutations in the *XPD* gene. Pyrimidine dimers are removed preferentially from the transcribed strand^{26,27}. Cockayne's syndrome cells are defective in preferential repair of the transcribed strand²⁸, which is consistent with the possible involvement of XPD in coupling transcription and excision repair. □

- Johnson, R. T. & Squires, S. *Mutat. Res.* **273**, 97–118 (1992).
- Reynolds, P., Higgins, D. R., Prakash, L. & Prakash, S. *Nucleic Acids Res.* **13**, 2357–2372 (1985).
- Nonet, M., Scafe, C., Sexton, J. & Young, R. *Molec. cell. Biol.* **7**, 1602–1611 (1987).
- Cornack, B. & Struhl, K. *Cell* **69**, 685–696 (1992).
- Herrick, D., Parker, R. & Jacobson, A. *Molec. cell. Biol.* **10**, 2269–2284 (1990).
- Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A. & Abelson, J. *Cell* **14**, 221–236 (1978).
- Wootner, M., Wade, P. A., Bonner, J. & Jaehning, J. A. *Molec. cell. Biol.* **11**, 4555–4560 (1991).
- Koleske, A. J., Buratowski, S., Nonet, M. & Young, R. A. *Cell* **69**, 883–894 (1992).
- Sung, P., Prakash, L., Matson, S. W. & Prakash, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8951–8955 (1987).
- Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. *Cell* **56**, 549–561 (1989).
- Zawel, L. & Reinberg, D. *Prog. nucl. Acid Res. molec. Biol.* **44**, 67–107 (1993).
- Bunick, D., Sandomeni, R., Ackerman, S. & Weinman, R. *Cell* **29**, 877–886 (1982).
- Sawadogo, M. & Roeder, R. G. *J. Biol. Chem.* **259**, 5321–5326 (1984).
- Serizawa, H., Conaway, J. W. & Conaway, R. C. *Nature* **363**, 371–374 (1993).
- Wang, W., Carey, M. & Gralla, J. D. *Science* **255**, 450–453 (1992).
- Feaver, W. J., Gileadi, O., Li, Y. & Kornberg, R. D. *Cell* **67**, 1223–1230 (1991).
- Conaway, R. C. & Conaway, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7356–7360 (1989).
- Schaeffer, L. et al. *Science* **260**, 58–63 (1993).
- Gulyas, K. D. & Donahue, T. F. *Cell* **69**, 1031–1042 (1992).
- Park, E., et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 11416–11420 (1992).
- Sung, P., Prakash, L., Weber, S. & Prakash, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6045–6049 (1987).
- Bailly, V., Sung, P., Prakash, L. & Prakash, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9712–9716 (1991).
- Lehmann, A. R. *Cancer Rev.* **7**, 82–103 (1987).
- Mellon, I., Spivak, G. & Hanawalt, P. C. *Cell* **51**, 241–249 (1987).
- Mellon, I. & Hanawalt, P. C. *Nature* **342**, 95–98 (1989).
- Venema, J. thesis, State Univ. Leiden, The Netherlands (1991).
- Madura, K. & Prakash, S. *J. Bact.* **166**, 914–923 (1986).
- Ogden, R. C., Lee, M.-C. & Knapp, G. *Nucleic Acids Res.* **12**, 9367–9382 (1984).

ACKNOWLEDGEMENTS. S.N.G. and H.Q. contributed equally to this work. We thank R. Schiestl for the *rad3 ts* strain, and J. Hopper, G. Sprague, Y. Surdin-Kerjan, R. Young and H. Zalkin for various plasmids. This work was supported by the US National Cancer Institute and Department of Energy.

Structural basis of superantigen action inferred from crystal structure of toxic-shock syndrome toxin-1

K. Ravi Acharya*, Edward F. Passalacqua*, E. Yvonne Jones†, Karl Harlos†, David I. Stuart†, Rossalyn D. Brehm‡ & Howard S. Tranter‡

* School of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK

† Oxford Centre for Molecular Sciences and Laboratory of Molecular Biophysics, Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK

‡ Division of Biologics, PHLS Centre for Applied Microbiology and Research, Porton Down, Salisbury SP4 0JG, UK

SUPERANTIGENS stimulate T cells bearing particular T-cell receptor $V\beta$ sequences^{1,2}, so they are extremely potent polyclonal T-cell mitogens. T-cell activation is preceded by binding of superantigens to class II major histocompatibility complex (MHC) molecules³. To further the structural characterization of these interactions, the crystal structure of a toxin associated with toxic-shock syndrome, TSST-1, which is a microbial superantigen, has been determined at 2.5 Å resolution. The N- and C-terminal domains of the structure both contain regions involved in MHC class II association; the C-terminal domain is also implicated in binding the T-cell receptor. Despite low sequence conservation, the TSST-1 topology is similar to the structure reported for the superantigen staphylococcal enterotoxin B⁴. But TSST-1 lacks several of the structural features highlighted as central to superantigen activity in the staphylococcal enterotoxin B and we therefore reappraise the structural basis of superantigen action.

Toxic-shock syndrome toxin-1 (TSST-1) is secreted by *Staphylococcus aureus* and is important in the genesis of human toxic-shock syndrome⁵. The protein has 194 residues and is

Received 2 August; accepted 29 November 1993.

- Prakash, S., Sung, P. & Prakash, L. A. *Rev. Genet.* **27**, 33–70 (1993).
- Weber, C. A., Salazar, E. P., Stewart, S. A. & Thompson, L. H. *EMBO J.* **9**, 1437–1447 (1990).

monomeric in solution (data not shown). The crystal structure determination is described in Table 1. The two domains of TSST-1 are packed tightly together to form a compact molecule of dimensions $\sim 45 \times 45 \times 58 \text{ \AA}^3$ (Fig. 1). After a short N-terminal portion, which spans the domain interface and includes an α -helix ($\alpha 1$), the polypeptide folds in a continuous 'roll' of β -strands ($\beta 1 - \beta 5$) to form the distinctive β -barrel of the N-terminal domain. This motif has also been noted in four other proteins of unrelated sequence⁶, including two bacterial toxins, the heat-labile toxin and verotoxin-1 from *Escherichia coli*. A piece of extended chain caps one end of the β -barrel in TSST-1, replacing the α -helix observed in the other structures. However, the topology and positioning of the structural elements is conserved (Fig. 1). In TSST-1, the concave $\beta 1\beta 2\beta 3$ face of this domain is distinguished by a prominent clustering of solvent-accessible hydrophobic residues (Fig. 2a). The C-terminal domain is composed of one particularly long α -helix ($\alpha 2$) with a highly twisted β -sheet ($\beta 6 - \beta 12$) cupped around it (Fig. 2b). As the N-terminal domain packs against the other face of $\alpha 2$, this helix serves as the 'backbone' of the molecule. The overall topology of the C-terminal domain is unique so far to the microbial superantigens, although it may be considered to be an elaborated form of the β -grasp motif⁶.

Figure 2a shows the buried residues of TSST-1 and, on the basis of the structure, possible linear epitopes can be suggested. Data for the location of biologically active sites on TSST-1 are related to the structure shown in Fig. 3a and b. Two results are specific for MHC class II binding. A peptide corresponding to residues 58–78, which reacts with neutralizing monoclonal antibodies against whole TSST-1 and demonstrates some of the immunoregulatory properties of the whole molecule, has been implicated in binding to MHC class II molecules⁷. Similarly, peptides 39–68 and 155–194 have been reported to compete with whole TSST-1 for MHC class II binding⁸. Thus, many of the data on specific binding to MHC class II molecules centre on residues 39–78 in the N-terminal β -barrel domain. In contrast, a C-terminal 12K fragment (M_r 12,000; residues 88–194, in effect the C-terminal domain) of TSST-1 has been reported to have mitogenic activity and to contain the epitope for antibody 8-5-7, which is able to neutralize the major clinical signs of toxic-shock syndrome^{9,10}. As superantigen/MHC class II binding is a prerequisite of subsequent T-cell antigen receptor (TCR) binding and mitogenic activity, this fragment must retain some MHC class II binding affinity in addition to the TCR binding site. Site-directed mutagenesis^{11,13} has implicated residues Tyr 115, Glu 132, His 135, Ile 140, His 141 and Tyr 144 in mitogenic activity. Tyrosine 115 is exposed to solvent in the $\beta 7\beta 8$ loop and as mutation to alanine abrogates antibody 8-5-7 binding, it appears to contribute directly to this key neutralizing epitope. Glutamic acid 132, His 135 and Ile 140 are all located on helix $\alpha 2$ and are exposed to solvent. It has been reported¹² that replacement of residue His 135 by alanine abolishes mitogenic activity, although the mutant protein is still recognized by antibody 8-5-7. Residues 141 and 144 are both buried, and mutation of these residues may well result in conformational changes influencing the nature of the molecular surface in the region of the $\alpha 2\beta 9$ loop (situated above Tyr 115 in the orientation of Fig. 3a, b). The extent of such changes is not sufficient to impinge on the 8-5-7 epitope for His 141 but may be for the deeply buried Tyr 144 (ref. 11). In total, the available data for the superantigenic activity of TSST-1 appear to cluster into two sites, one (site A) involving the concave face of the N-terminal β -barrel, the other (site B) containing $\alpha 2$ and the $\beta 7\beta 8$ and $\alpha 2\beta 9$ loops in the C-terminal domain. Site A is associated with MHC class II binding. But as the C-terminal domain alone⁹, and a portion of the polypeptide chain between residues 155 and 194 in particular, have also been implicated in MHC binding⁸, we predict that the residues between 170 and 180, where the chain packs closely against the concave face of the N-terminal β -barrel (Fig. 3a, b), form part of site A. Data for site B do not

TABLE 1 Statistics for crystallographic structure determination

Diffraction data					
Dataset	$d_{\min}$ (Å)	No. of measurements	Unique reflections (% complete)	R_{merge} (all data)	MFID
Native (IP Lab 1)	3.5	20,717	11,231 (94)	7.5	—
Native (IP Lab 2)	2.5	49,530	29,155 (87)	7.5	—
Native (IP Lab 1 + 2)	2.5	70,247	30,007 (89)	8.6	—
Uranyl acetate	3.5	11,104	9,186 (77)	8.4	29.9
Uranyl fluoride	3.5	13,938	10,536 (88)	4.5	17.1
MIR phasing					
Derivative	No. of sites		R_c to 6.0 Å (3.5 Å)	Phasing power to 6.0 Å (3.5 Å)	
Uranyl acetate (10 mM)	5		0.55	1.9	
Uranyl fluoride (10 mM)	3		0.49 (0.56)	2.1 (1.7)	
Solvent flattening					
$d_{\min}$ (Å)	Pre-Wang FM	% Solvent	Wang R factor	Phase change	Wang FM
6.0	0.60	65	26.3	41.3°	0.78
3.5	—	65 (55)	22.6 (18.7)	50.5° (48.3)	0.77 (0.80)

TSST-1 was purified and crystallized as described²⁴. The crystals belong to orthorhombic space group C22₁ (cell dimensions $a = 108.6$, $b = 177.6$, $c = 97.5$ Å) with 3 molecules per asymmetric unit, but possess pseudo-hexagonal symmetry space group P6₃22 (cell dimensions $a = 104.1$, $c = 97.5$ Å; with 1 molecule per asymmetric unit). Diffraction data from native crystals (3.5 Å and 2.5 Å resolution) and from two heavy-atom derivatives (3.5 Å resolution) were collected using an MAR-research imaging-plate system with CuK α X-rays generated by a Rigaku RU200 rotating anode with a graphite monochromator. Data were processed using the MOSFLM package (A. Leslie) ($R_{\text{merge}} = \sum (|I_{\text{obs}} - \bar{I}|) / \sum \bar{I}$; MFID, mean fractional isomorphous difference). Heavy-atom sites were identified using the automatic Patterson search program GROPAT²⁵ and difference Fourier analyses. Heavy-atom parameters were refined and MIR phases calculated using the programs REFIN and PHASE (P. Shaw, unpublished) respectively (R_c : Cullis R -factor for centric reflections; phasing power: mean value of the heavy-atom structure factor amplitudes divided by the residual lack-of-closure error). The MIR phases were improved by solvent flattening (65% solvent) to give a 3.5 Å electron density map from which we established the relationship between the non-crystallographically related molecules. The pseudo-symmetry arises because the orthorhombic asymmetric unit contains one half of a crystallographic dimer and one non-crystallographic dimer arranged on a pseudo-hexagonal net (FM, mean figure of merit; Wang R factor, crystallographic R factor between observed structure factor amplitudes and those calculated from the solvent-flattened map; phase change, difference between MIR phases and phases from the solvent flattened map). *Ab initio* phase extension to 2.5 Å was achieved using the GAP program suite (D.I.S. and J. Grimes, unpublished) by 3-fold real-space averaging, which resulted in an R -factor and correlation coefficient of 0.31 and 0.91, respectively, between observed data and data obtained by inversion of the averaged solvent-flattened map. Owing to the 3-fold averaging, the 2.5 Å resolution electron density map was quite clear. Initially a skeletal model for one monomer (all 194 residues) was fitted to the 2.5 Å electron density map (program FRODO²⁶) and the appropriate amino-acid sequence incorporated using the program CALPHA (R. Esnouf, unpublished). The structure was refined by simulated annealing with strict non-crystallographic symmetry constraints (program XPLOR²⁷). In the final round, the non-crystallographic constraints were relaxed and the 3 molecules were individually checked and refined with non-crystallographic restraints. The current model, which includes 4,677 non-hydrogen atoms (for 3 molecules), has a crystallographic R factor of 0.213 for all data to 2.5 Å (15.6% for $F \geq 3\sigma$). No bound waters have been modelled, although a correction for the bulk solvent has been made. The root-mean-square (r.m.s.) deviation from ideal values for bond lengths is 0.011 Å and for bond angles is 2.68°. The r.m.s. deviation of atomic coordinates for all atoms between monomer pairs 1–2, 1–3 and 2–3 are 0.13, 0.15, 0.15 Å, respectively. With the exception of Ser 86 ($\phi = 61^\circ$, $\psi = -157^\circ$), main-chain torsion angles for all residues lie within allowed regions of the Ramachandran plot. Atomic coordinates will be deposited in the Brookhaven Protein Data Bank.

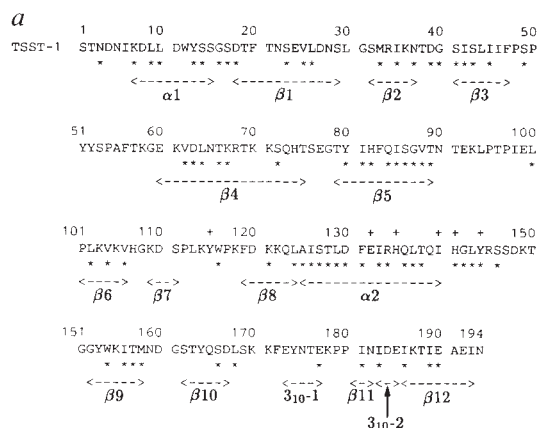
distinguish between MHC class II or TCR binding, but we suggest that its position (diametrically opposed on the TSST-1 molecule to site A) makes it the likely site of TCR interaction.

TSST-1 shares little amino-acid sequence identity ($\sim 24\%$) with the staphylococcal enterotoxins (SEs: SEA–SEE) but shares similar biological functions. A comparison on the basis of secondary structure assignments is shown for TSST-1 and SEB in Fig. 2b. The smaller protein, TSST-1, has the same core topology as SEB but is extensively truncated in several loop

regions, including some that incorporate secondary structure elements. Thus there are no structural equivalents in TSST-1 for much of the SEB N-terminal region, including SEB $\alpha 1$. Similarly, helices $\alpha 3$ and $\alpha 5$, situated in loops between $\beta 3\beta 4$ and $\beta 10\beta 11$ in SEB, are respectively lost and truncated to a single turn of 3_{10} helix in TSST-1, whereas the $\beta 2\beta 3$ and $\beta 4\beta 5$ loops are markedly shorter. This latter loop is often referred to as the 'cysteine loop' of the SEs, and residues in this region have been implicated in SE emetic activity¹⁴. The lack of emetic activity in TSST-1 is consistent with the marked difference we observe between TSST-1 and SEB structures in this region. The many alterations in TSST-1 as compared with SEB, particularly in the N-terminal region and SEB $\alpha 5$, both of which are situated at the domain interface, also have considerable bearing on the superantigenic sites, as predicted from inspection of the SEB structure⁴. Inhibition of intact SEA by synthetic peptides has implicated the N-terminal domain of SEA in MHC class II binding^{15,16}. Mutagenesis¹⁷ of the solvent-accessible Asn 23 of SEB implicates $\alpha 2$ of that molecule in TCR binding. Data from the same study for the exposed hydrophobic residue Phe 44 (in SEB) associate the $\beta 1\beta 2$ turn with MHC class II binding. This study also demonstrated $V\beta$ -specific binding effects at the tip of the $\beta 2\beta 3$ loop and some MHC class II binding dependency on $\alpha 1$ (SEB); these areas have no structural equivalents in TSST-1. In SEA, mutants that affect $V\beta$ specificity¹⁸ map onto residues just preceding the start of $\alpha 5$ in SEB; again, there are no structural equivalents in TSST-1. With the exception of Asn 23 (SEB), mutations in SEA and SEB reported to affect TCR binding abrogate binding to only some of the $V\beta$ elements known to recognize the native protein.

Many of the detailed elements of both TCR and MHC binding sites appear to be altered between TSST-1 and the SE family, as represented by SEB. Overall, the evidence for MHC class II binding to these microbial superantigens focuses onto the front ($\alpha 5$ in SEB) face of the molecule (Fig. 3a, b), which includes the concave surface of the $\beta 1\beta 2\beta 3$ sheet in the N-terminal domain. It is therefore of particular interest that this surface, delineated by the $\beta 1\beta 2$, $\beta 3\beta 4$ and $\beta 4\beta 5$ loops, is the site of oligonucleotide or oligosaccharide binding in the structurally analogous domains of other proteins⁶. The specific site of TCR binding involves several amino acids on the top and back ($\alpha 4$ in SEB) face of the molecule (Fig. 3a, b) and appears to vary between different superantigens and/or different TCR $V\beta$ variants.

Any discussion of superantigen binding must consider the structure of the MHC class II and TCR molecules. MHC class II/superantigen binding is believed to involve regions to the



b

SEB		TSST-1	
Secondary structure	Residues	Secondary structure	Residues
$\alpha 1$	12-17	-	-
$\alpha 2$	21-29	$\alpha 1$	7-15
$\beta 1$	33-39	$\beta 1$	18-29
$\beta 2$	48-52	$\beta 2$	32-37
$\beta 3$	63-68	$\beta 3$	41-47
$\alpha 3$	70-78	-	-
$\beta 4$	81-89	$\beta 4$	60-75
$\beta 5$	112-120	$\beta 5$	79-89
$\beta 6$	127-138	$\beta 6$	101-106
$\beta 7$	141-151	$\beta 7$	109-111
$\beta 8$	154-156	$\beta 8$	119-124
$\alpha 4$	157-172	$\alpha 2$	125-140
$\beta 9$	182-190	$\beta 9$	152-158
$\beta 10$	195-200	$\beta 10$	161-166
$\alpha 5$	210-217	3_{10-1}	173-177
$\beta 11$	222-225	$\beta 11$	181-182
-	-	3_{10-2}	183-185
$\beta 12$	229-236	$\beta 12$	186-193

FIG. 2 **a**, Amino-acid sequence of TSST-1. Positions of the secondary structural elements in TSST-1 (based on program DSSP) are indicated, and solvent-inaccessible residues (less than 20 Å² of exposed surface) are indicated by asterisks. Residues implicated in mitogenic activity (based on site-directed mutagenesis) are represented by plus signs. **b**, Comparison of the secondary structural elements for SEB⁴ and TSST-1.

side of the antigen-binding cleft^{19,20}. Superantigen binding to the TCR is thought to be to the solvent-exposed face of a β -sheet (the BED β -sheet) in the $V\beta$ component of the TCR $V\alpha V\beta$ dimer^{21,22}. Combination of these observations with a model for the MHC/TCR antigen recognition complex enables us to represent superantigen action schematically (Fig. 3c)²³. Figure 3d shows that these proposed sites of interaction between TSST-1 and the MHC and TCR molecules are entirely consistent with such a model. In the absence of high-resolution evidence for the structure of the ternary complex, such models may prove helpful for understanding the microbial superantigen mechanism of pathogenicity and enable therapeutic compounds to be engineered that target appropriate epitopes derived from the molecular structures of these toxins. □

Received 8 June; accepted 2 November 1993.

- White, J. et al. *Cell* **56**, 27-35 (1989).
- Janeway, C. A. et al. *Immun. Rev.* **107**, 61-88 (1989).
- Fleischer, B. & Schrezenmeier, H. *J. exp. Med.* **167**, 1697-1707 (1988).
- Swaminathan, S. et al. *Nature* **359**, 801-806 (1992).
- Marrack, P. & Kappler, J. *Science* **248**, 705-711 (1990).
- Murzin, A. G. *Nature* **360**, 635 (1992).
- Edwin, C. et al. *J. infect. Dis.* **163**, 524-529 (1991).
- Soos, J. M. et al. *Biochem. biophys. Res. Commun.* **191**, 1211-1217 (1993).
- Edwin, C. & Kass, E. H. *Infect. Immun.* **57**, 2230-2236 (1989).
- Bonventre, P. F. et al. *Infect. Immun.* **56**, 135-141 (1988).

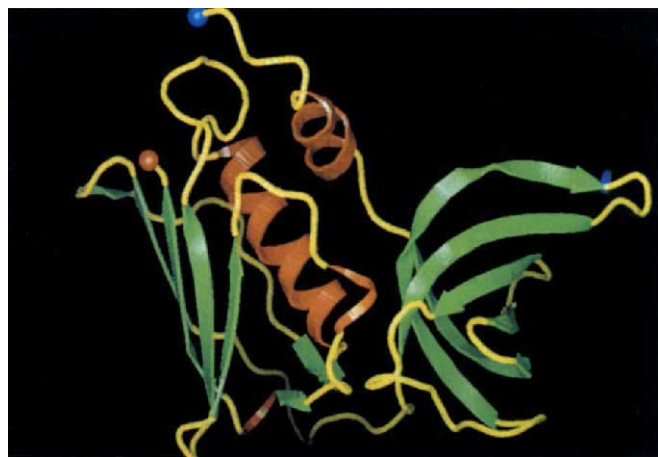


FIG. 1 The polypeptide fold for TSST-1. Helices are coloured red (3_{10} helices being drawn more thinly than α -helices), β -strands green and loops yellow. Blue and red balls represent N and C termini, respectively.

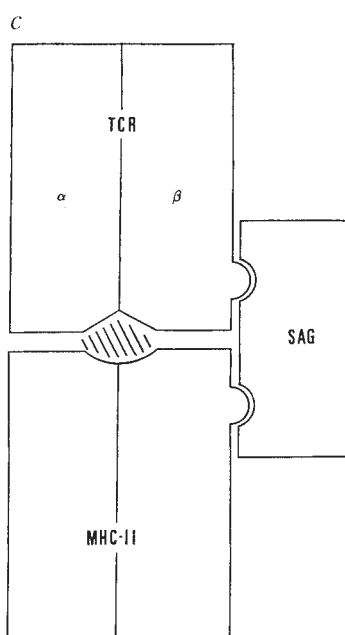
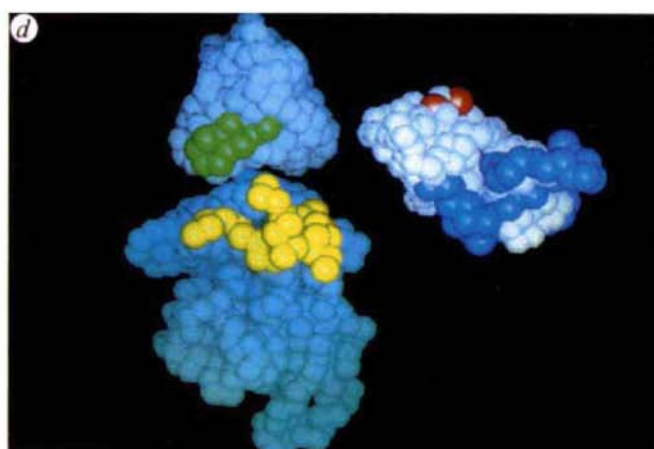


FIG. 3 MHC class II and TCR binding to TSST-1. *a* and *b*, Residues on TSST-1 implicated in these interactions. In *a*, the molecule is viewed as in Fig. 1; in *b*, the molecule has been rotated by 180° about a vertical axis. The regions implicated by peptide data in MHC class II binding are coloured yellow (residues 39–78) and green (residues 155–194)⁸, the rest of the molecule in blue. Residues implicated in mitogenic activity (based on mutational analysis^{11–13}) are highlighted in red. The cyan and green balls represent the N and C termini, respectively. *c*, Standard schematic representation of superantigen (SAG) function (cross-hatched region represents the MHC-bound peptide). *d*, Possible model for the interaction of TSST-1 with MHC and TCR molecules. The interacting faces of both the MHC–TCR complex and the TSST-1 are revealed by (opposed) 90° rotations of each about a vertical axis. The mid-blue, pale blue and white space-filled objects represent the MHC molecule, the TCR molecules ($V\alpha$ and $V\beta$ domains), and the TSST-1 molecule, respectively. The MHC/TCR interactions are modelled after ref. 28 using MHC class I and immunoglobulin light-chain dimer as representative coordinates. The yellow area on the MHC molecule maps the location of the regions $\alpha 40$ to $\alpha 50$ and $\beta 81$ to $\beta 90$. Such an area is consistent with available data on superantigen binding^{19,20} but places particular emphasis on residues around one end of the peptide-binding cleft. The green region of the TCR molecule marks the DE loop on the external β -sheet surface of the $V\beta$ domain²⁹. Two opposed sites on the TSST-1 molecule, site A (represented by residues 39–78 and 170–180) and site B (indicated by residues 115, 132, 135 and 140), which we propose interact with the MHC class II and TCR molecules, are highlighted in dark blue and red, respectively. This model merely serves to relate gross structural features; data are drawn from studies using several different superantigens so details of binding are expected to vary. However, to within the low resolution of this exercise, such superantigen interactions are compatible with the standard models of MHC/TCR complex formation. Our model involves superantigen binding at a site on the end of the MHC class II molecule peptide-binding cleft opposite the region ($\beta 49$ to $\beta 55$) recently implicated in MHC class II dimerization³⁰. The side of the MHC class II molecule proposed³⁰ as the binding site of CD4 also remains accessible in our model. Thus, this form of superantigen binding could simply act as an additional stabilizing component in the multimolecular aggregations of MHC class II, TCR and CD4 molecules proposed by Brown *et al.*³⁰.

11. Blanco, L. *et al.* *Infect. Immun.* **58**, 3020–3028 (1990).
12. Bonventre, P. F. *et al.* *Infect. Immun.* **61**, 793–799 (1993).
13. Murray, D. L. *et al.* *Abstr. 93rd gen. Mtg Am. Soc. Microbiol.*, Atlanta, Georgia, 64 (1993).
14. Spero, L., Johnson-Winegar, A. & Schmidt, J. J. in *Bacterial Toxins, Handbook of Natural Toxins* Vol. 4 (eds Hardegree M. C. & Tu A. T.) 131–163 (1988).
15. Griggs, N. D. *et al.* *J. Immun.* **148**, 2516–2521 (1992).
16. Pontzer, C. H., Russell, J. K. & Johnson, H. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 125–128 (1991).

17. Kappler, J. W. *et al.* *J. exp. Med.* **175**, 387–396 (1992).
18. Irwin, M. J. *et al.* *Nature* **359**, 841–843 (1992).
19. Braunstein, N. S. *et al.* *J. exp. Med.* **175**, 1301–1305 (1992).
20. Panina-Bordignon, P. *et al.* *J. exp. Med.* **176**, 1779–1784 (1992).
21. Pullen, A. M. *et al.* *Cell* **61**, 1365–1374 (1990).
22. Pullen, A. M. *et al.* *J. exp. Med.* **173**, 1183–1192 (1991).
23. Woodland, D. L. & Blackman, M. A. *Immun. Today* **14**, 208–212 (1993).
24. Passalacqua, E. F. *et al.* *J. molec. Biol.* **228**, 983–986 (1992).
25. Jones, E. Y., Walker, N. P. C. & Stuart, D. I. *Acta crystallogr.* **A47**, 753–770 (1991).
26. Jones, A. *Meth. Enzym.* **115**, 157–171 (1985).
27. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
28. Bjorkman, P. J. & Davis, M. M. *Cold Spring Harb. Symp. quant. Biol.* **LIV**, 365–373 (1989).
29. Choi, Y. *et al.* *Nature* **346**, 471–473 (1990).
30. Brown, J. H. *et al.* *Nature* **364**, 33–39 (1993).

ACKNOWLEDGEMENTS. We thank A. R. Rees and P. Hambleton for support and encouragement, V. Subramanian, T. Edwards, H. Reyburn and S. Davis for discussion, P. Artymuk for the database search, and the SERC for a studentship to E.F.P.; D.I.S. is a member and E.Y.J. an associate member of the Oxford Centre for Molecular Sciences supported by the SERC and MRC. E.Y.J. is supported by a Royal Society University Research Fellowship and K.H. by the OCMS.

Snapshots of the 'molten globule'

The first issue of this new monthly journal provides a variety of papers for structural biologists (and others) to get their teeth into, among them two dealing with the intriguing matter of protein folding.

GIVEN the number of possible conformations that a polypeptide chain can adopt it is remarkable that proteins fold to a unique tertiary structure at all, let alone achieve the feat in the order of seconds or minutes. So how do they do it? Reports in the January issue of *Nature Structural Biology*, by Feng *et al.*¹ and Redfield *et al.*², describe the structures of two partially unfolded proteins, apocytochrome b_{562} and interleukin-4, which provide hints about what is happening towards the end of the folding pathway.

Protein folding is clearly a directed process. The challenge is to divine the physico-chemical properties of proteins *in vitro*, and the role of the *in vivo* co-factors that aid and abet the transformation of a linear sequence of amino acids into a functioning, paid-up protein. One of the main difficulties in studying protein folding *in vitro* is the transient nature of the kinetic intermediates on the pathway. In consequence these species are for the most part only ever present in low stoichiometric amounts, making analysis of their structure extremely tricky. One solution is to look at intermediates between the folded and unfolded state that, under certain conditions, occur at equilibrium in sufficient quantity that they can be studied at leisure. One such species is the molten globule, which has properties similar to those of the kinetic intermediates.

Relatively stable, partially folded proteins can be generated by a number of means (variation of the solvent conditions, removal of a ligand or prosthetic group and so on). In general, a partially folded protein is considered to be a molten globule if it has the following properties: a well-defined secondary structure, some of which may be in a native-like conformation; none of the specific tertiary interactions that define the native state; and an expanded structure relative to the folded protein.

Cytochrome b_{562} is a haem-containing protein that consists of four helices with a binding pocket in which the large haem prosthetic group nestles (the holoprotein). Removal of the haem results in partial unfolding of the protein, which nonetheless maintains much of its structure under near-physiological conditions (the apoprotein).

The solution structure of the apoprotein, as determined by Feng *et al.*, is remarkably similar to the structure of the holoprotein. But although three of the four helices in the holoprotein are fully represented in the apoprotein, distortion of the C terminus of helix IV and the misalignment of helix I and II result in poor packing interfaces between these secondary structure elements.

The various derangements of the apoprotein result in the exposure of the haem-binding pocket to the surrounding solvent and the complete solvation of the amino-acid residues, Met 7 and His 107, that provide axial ligands to the haem. The residues that contact the haem group, and other residues that form the core of the protein, are not repacked in the apoprotein and remain exposed to solvent. The large cavern thus formed in the apoprotein is huge, and able to accommodate up to fifty water molecules.

But can the apoprotein be said to be a molten globule? Although Feng *et al.* wisely hesitate about claiming that it should have full molten-globule status, it is clear that the apoprotein has some of the appropriate features. But argument over that question is in some senses a side issue, for the structure is of interest from another point of view. As the authors point out, the NMR constraints were obtained under near-physiological conditions suggesting that the structure of the apoprotein represents a folding intermediate late on the pathway in haem protein assembly.

Redfield and colleagues' structure analysis is of interleukin-4 (IL-4), a four-helix-bundle protein, at low pH. Both the loss of the near-ultraviolet circular dichroism (CD) spectrum and the enhanced fluorescence of the hydrophobic dye 1-anilinonaphthalene-8-sulphonic acid (ANS) in the presence of IL-4 under acid conditions are characteristic of the canonical molten globule state.

Despite this, comparison of the protein in roughly neutral and acidic conditions again reveals that there is relatively little

change in the NMR spectra for most of the amino-acid residues, suggesting that the structure at low pH is similar to that of the native form of the protein. Even so, more than one third of the residues are in regions of significant disorder. The four helices, which form the core of the protein, are for the most part highly ordered. It is the sequences which connect the helices that are substantially disordered.

Determination of order parameters, which provide an indication of the rigidity of the main chain of the protein, indicate that the region around the amino terminus of the C helix undergoes a local unfolding transition at low pH. So it would seem that the 'molten globule-like' ANS and near-ultraviolet CD characteristics are a consequence of increased disorder of localized portions of the structure, rather than the formation of a globally disordered state.

The structures of apocytochrome b_{562} and IL-4 are both highly ordered compared to some of the more disordered molten globules characterized so far. Nonetheless both proteins have some of the features of the molten globule state and, as Redfield *et al.* suggest for IL-4, both might reasonably be considered as 'highly ordered molten globules'.

Clearly, the label of molten globule can encompass a range of degrees of unfolding of a polypeptide chain. Those that lack most residual tertiary structure and are highly disorganized are probably related to kinetic intermediates formed early in the folding pathway. The 'highly ordered molten globules' characterized by Feng *et al.* and Redfield *et al.* are likely to be similar to the types of structures occurring late in the protein folding process.

The burning question is this — what happens *in vivo*? The various proteins that assist in folding (protein disulphide isomerase, chaperones and so on) do not seem to impart structural information to the polypeptide chains with which they interact. So one would hope that the lessons learnt *in vitro* will illuminate protein folding in the cell.

Guy Riddihough

Guy Riddihough is Editor of Nature Structural Biology.

Also in *Nature Structural Biology* in January: substrate-ribozyme interactions; using isosteric base analogues as a means to map the interactions between the *trp* repressor and its cognate operator sequence; intact-cleaved human antithrombin III complex as a model for serpin-proteinase interactions; structure of the toxin from *Shigella dysenteriae* responsible for the severe form of dysentery in humans.

1. Feng, Y., Silgar, S.G. & Wand, A.J. *Nature struct. Biol.* **1**, 30–35 (1994).
2. Redfield, C., Smith, R.A.G. & Dobson, C.M. *Nature struct. Biol.* **1**, 23–29 (1994).

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963 65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.